



Review

Plant species identification using digital morphometrics: A review

James S. Cope^a, David Corney^{b,*}, Jonathan Y. Clark^b, Paolo Remagnino^a, Paul Wilkin^c^a Digital Imaging Research Centre, Kingston University, London, UK^b Department of Computing, University of Surrey, Guildford, Surrey, UK^c Royal Botanic Gardens, Kew, Richmond, Surrey, UK

ARTICLE INFO

Keywords:

Morphometrics
Shape analysis
Image processing
Plant science
Leaf
Flower
Taxonomy

ABSTRACT

Plants are of fundamental importance to life on Earth. The shapes of leaves, petals and whole plants are of great significance to plant science, as they can help to distinguish between different species, to measure plant health, and even to model climate change. The growing interest in biodiversity and the increasing availability of digital images combine to make this topic timely. The global shortage of expert taxonomists further increases the demand for software tools that can recognize and characterize plants from images. A robust automated species identification system would allow people with only limited botanical training and expertise to carry out valuable field work.

We review the main computational, morphometric and image processing methods that have been used in recent years to analyze images of plants, introducing readers to relevant botanical concepts along the way. We discuss the measurement of leaf outlines, flower shape, vein structures and leaf textures, and describe a wide range of analytical methods in use. We also discuss a number of systems that apply this research, including prototypes of hand-held digital field guides and various robotic systems used in agriculture. We conclude with a discussion of ongoing work and outstanding problems in the area.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Plants form a fundamental part of life on Earth, providing us with breathable oxygen, food, fuel, medicine and more besides. Plants also help to regulate the climate, provide habitats and food for insects and other animals and provide a natural way to regulate flooding. A good understanding of plants is necessary to improve agricultural productivity and sustainability, to discover new pharmaceuticals, to plan for and mitigate the worst effects of climate change, and to come to a better understanding of life as a whole.

With a growing human population and a changing climate, there is an increasing threat to many ecosystems. It is therefore becoming increasingly important to identify new or rare species and to measure their geographical extent as part of wider biodiversity projects. Estimates of numbers of species of flowering plants (or angiosperms) vary from about 220,000 (Scotland & Wortley, 2003) to 420,000 (Govaerts, 2001).

The traditional approach to identifying species and their relationships is to train taxonomists who can examine specimens and assign taxonomic labels to them. However, there is a shortage of such skilled subject matter experts (a problem known as the “taxonomic impediment” e.g. Carvalho et al., 2007), as well as a

limit on financial resources. Furthermore, an expert on one species or family may be unfamiliar with another. This has led to an increasing interest in automating the process of species identification and related tasks. The development and ubiquity of relevant technologies, such as digital cameras and portable computers has brought these ideas closer to reality; it has been claimed that now is the “time to automate identification” (MacLeod, Benfield, & Culverhouse, 2010), and not just of plants. Arguing that we need to train more expert taxonomists, while also embracing new technologies, Quentin Wheeler writes that “[d]igital images are to morphological knowledge what the Gutenberg Press was to the written word” (Wheeler, 2004).

Botanists collect specimens of plants and preserve them in archives in herbaria. For example, the herbarium at the Royal Botanic Gardens, Kew in London houses over 7 million dried specimens,¹ some of which are more than 200 years old. These are annotated and sorted using the expert knowledge of the botanists, subject to revision over time. Herbarium collections can therefore be seen as major, structured repositories of expert knowledge. In order to improve access, these collections are increasingly being digitized to form databases with images that are annotated with species’ names, collectors’ names, dates, locations and so on. Other significant sources of knowledge include flora, taxonomic keys and monographs (see Table 1 for definitions).

* Corresponding author.

E-mail addresses: J.Cope@kingston.ac.uk (J.S. Cope), D.Corney@surrey.ac.uk (D. Corney), J.Y.Clark@surrey.ac.uk (J.Y. Clark), P.Remagnino@kingston.ac.uk (P. Remagnino), P.Wilkin@kew.org (P. Wilkin).¹ <http://apps.kew.org/herbcat/navigator.do>.

Table 1

Some botanical terminology. Note that some terms have different meanings in plant science compared to computer science or statistics. See also Fig. 2 for terms relating to the anatomy of a leaf.

Identification	Recognition of the identity of an organism. (Synonymous with <i>classification</i> in computer science and statistics.)
Classification	Grouping items based on similarity. (Synonymous with <i>cluster analysis</i> or <i>segmentation</i> in computer science and statistics.)
Nomenclature	Assigning names to organisms
Taxonomy	Identification, formal description and naming of organisms
Taxon	Group of organisms assumed to be a unit; e.g. a species
Taxonomic rank	Relative position in taxonomic hierarchy; e.g. “species”, “family”
Dichotomous key	A binary tree that allows a user to identify members of a taxon through a series of questions
Flora	A book describing plant life in a particular geographic region
Monograph	A book providing a (near) complete description of a particular taxon, typically a genus
Taxonomic key	Structured series of questions used to identify specimens
Herbarium	A reference collection of preserved plant specimens.
Systematics	The taxonomic study of evolutionary origins and environmental adaptations
Cladistics	The study of the pathways of evolution, with the aim of identifying ancestor–descendant relationships
Phenetics	The study of relationships between organisms defined by the degree of physical similarity between them
Homology	The similarity of a structure in different organisms resulting from shared ancestry

In many cases, species (or higher taxa such as genera or families) can be distinguished by characters derived from their leaf or flower shape, or their branching structure. Shape is of course important in many other disciplines, and morphometric techniques are applied in structurally-based research in zoology, geology, archaeology and medicine, although these are beyond the scope of this review.

Morphometrics, the study of shape, has been applied to plants and their organs for many years. Leaves are readily apparent structures on many plants, and they are available for examination for much of the year in deciduous or annual plants or year-round in evergreen perennials, unlike more transient reproductive organs. As such, leaf characters, including those involving shape, have been used extensively in traditional text-based taxonomic keys for plant identification since the beginnings of botany. Examples of such studies include those on *Tilia* (Schneider, 1912), *Ulmus* (Melville, 1937, 1939) and *Betula* (Natho, 1959), but there are many more. To use such a key, which has been compiled by an expert in the group in question, the user makes a series of choices between contrasting statements, finally reaching a species name. Even given such a key, the user must make a number of judgements that require specific botanical knowledge, so these cannot be used naively. Further details on taxonomic keys can be found in Stace (Stace, 1992).

In recent years, high quality digital cameras have become ubiquitous, increasing interest in creating hand-held field guides. These are prototypes built around smart phones or personal digital assistants (PDAs) that are designed to allow a user in the field to photograph a specimen of interest and instantly receive information about it, such as the likely species name (see also Section 3). One advantage of such systems is that they require little infrastructure at the point of use, so can be used even in the least developed and most remote parts of the world. However, the scope of such systems is currently very limited, restricting their practical use.

A second consequence of cheap digital cameras and scanners is the creation of vast databases of plant images. For example, the Royal Botanic Gardens, Kew provides a digital catalogue of over 200,000 high-resolution images, with more being added continuously as part of an ongoing digitization project. We maintain an annotated list of botanical image sets online,² describing various publicly available sets, including images of single-leaves, herbarium specimens, and whole plants.

1.1. Challenges in botanical morphometrics

Although morphometrics and image processing are well-established and broad disciplines, botanical morphometrics presents

some specific challenges. Here, we discuss some of these, including specimen deformations, unclear class boundaries, feature selection and terminology.

Leaves and flowers are non-rigid objects, leading to a variety of deformations. Many leaves have a three-dimensional nature, which increases the difficulty of producing good quality leaf images and also results in the loss of useful structure information. Archived specimens may also be damaged as they are dried and pressed, but even live specimens may have insect, disease or mechanical damage. Automated systems must be robust to such deformations, making soft computing and robust statistics highly attractive.

One source of confusion when botanists and computer scientists collaborate concerns terms such as “classify” and “cluster”. In taxonomy, “classification” may be defined as the process of grouping individuals based on similarity, in order to define taxa such as species or genera (Stuessy, 2006). “Identification” is then the process of deciding to which of a number of pre-defined taxa a particular individual belongs. In computer science by contrast, “classification” refers to the assigning of an individual example to one of a finite number of discrete categories, whereas “clustering” refers to the discovery of groups within a set of individuals, based on similarities (Bishop, 2007, p.3). Care must be taken when using such terms to avoid confusion.

Any system that is concerned with distinguishing between different groups of plants must be aware of the large intra-class, and small inter-class variation that is typical of botanical samples (see Fig. 1). A number of classifiers have been developed that identify the species of a specimen from a digital image, as we discuss throughout this paper, and these must be robust to this challenge. Similar issues apply to the tasks of discovering how many groups exist in a set of examples, and what the class boundaries are. See Figs. 3 and 7 for further examples of the variety of leaf shapes found.

Distinguishing between a large number of groups is inherently more complex than distinguishing between just a few, and typically requires far more data to achieve satisfactory performance. Even if a study is restricted to a single genus, it may contain many species, each of which will encompass variation between its constituent populations. The flowering plant genus *Dioscorea*, for example, contains over 600 species (Govaerts, Wilkin, Raz, & Téllez-Valdés, 2010), so even single-genus studies can be very challenging. On a related note, the shape of leaves may vary continuously or discretely along a single stem as the leaves develop (known as leaf heteroblasty), which can further confound shape analysis unless careful attention is paid to sources of the specimens.

Different features are often needed to distinguish different categories of plant. For example, whilst leaf shape may be sufficient to

² <http://www.computing.surrey.ac.uk/morphidas/ImageSets.html>.



Fig. 1. Variation in leaves taken from a single specimen of *Quercus nigra*.

distinguish between some species, other species may have very similar leaf shapes to each other, but have different coloured leaves. No single feature, or kind of feature, may be sufficient to separate all the categories, making feature selection a challenging problem.

Besides these botany-specific issues, more general image processing issues, such as the ambiguity caused by unknown illumination, pose etc., remain potentially problematic. This is especially true in field conditions with less control over the image capturing process.

1.2. Outline

The scope of this paper is focussed on approaches to plant species identification using digital images combined with domain knowledge. We aim to review current methods and applications, to highlight parallel streams of research and to motivate greater efforts to solve a range of important, timely and practical problems. This paper provides a thorough introduction to the main issues in this large and important field. We assume some basic familiarity with computing issues and terminology, but aim to introduce the reader to a range of concepts and issues in botany throughout the text, some of which are highlighted in Table 1.

The rest of this paper is organized as follows. In Section 2, we review a range of methods that have been applied to analyzing leaf shape, venation, leaf margin features, leaf texture and so on. We then discuss a number of systems designed for practical use in the field in Section 3, before a concluding discussion.

2. Leaf analysis methods

There are many aspects of a plant's structure and appearance that are used by expert botanists in plant morphological research. The most useful of these features are usually the two-dimensional outline shape of a leaf or petal (Section 2.1), the structure of the vein network (Section 2.2), and the characters of the leaf margin (Section 2.3). Of these, the outline shape has received by far the most attention when applying computational techniques to botanical image processing.

As well as being useful in their own right, the automatic extraction of such features is also an essential component of larger systems for species identification and related tasks. All forms of shape analysis can be viewed as methods to represent the implicit data of a raw image in a more useful form for subsequent processing.

2.1. Leaf shape

There are several reasons underlying the focus on leaf shape. Firstly, the shape has perhaps the most discriminative power. Although leaves from the same plant may differ in detail, it is often

the case that different species have characteristic leaf shapes, and these have often been used by botanists to identify species. Whereas differences in margin character or vein structure may be fairly subtle, shape differences are often more obvious, even to the non-expert. In many cases, leaf size is largely determined by the environment, while *shape* is more heritable. Secondly, this is the easiest aspect to automatically extract. If a leaf is imaged against a plain black or white background, then simple threshold techniques can be used to separate the leaf from the background, and the outline can then be found by simply isolating those pixels of the leaf that border the background. Thirdly, there are numerous existing morphometric techniques which can be applied to leaf shape that have already proven their worth for other biological problems and which may already be familiar to many botanists. Finally, the gross structure of a leaf may be preserved even if the leaf specimen is damaged, possibly through age. For example, many dried leaves turn brown, so colour is not usually a useful feature by itself. Note also that many of the shape-based methods discussed here have also been applied to petal, sepal or whole flower shape, as discussed in Section 2.6.

Fig. 2 shows some of the main features of leaves with their corresponding botanical terms, while Fig. 3 illustrates some of the variety of leaf shapes found.

We now discuss a number of approaches to leaf shape analysis, including Fourier analysis, contour signatures, landmark analysis, shape features, fractal dimensions and texture analysis.

2.1.1. Elliptic Fourier descriptors

One of the most common shape analysis technique applied to leaves is elliptic Fourier descriptors (EFDs, or elliptic Fourier analysis; EFA) (Kuhl & Giardina, 1982). Here, leaf shape is analyzed in the frequency domain, rather than the spatial domain. A set number of Fourier harmonics are calculated for the outline, each of

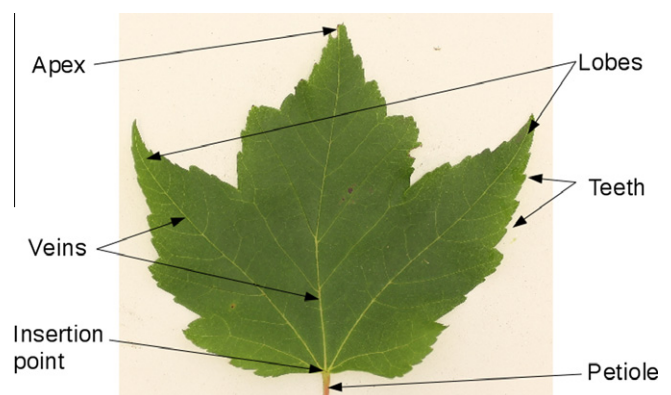


Fig. 2. The main features of a typical leaf.

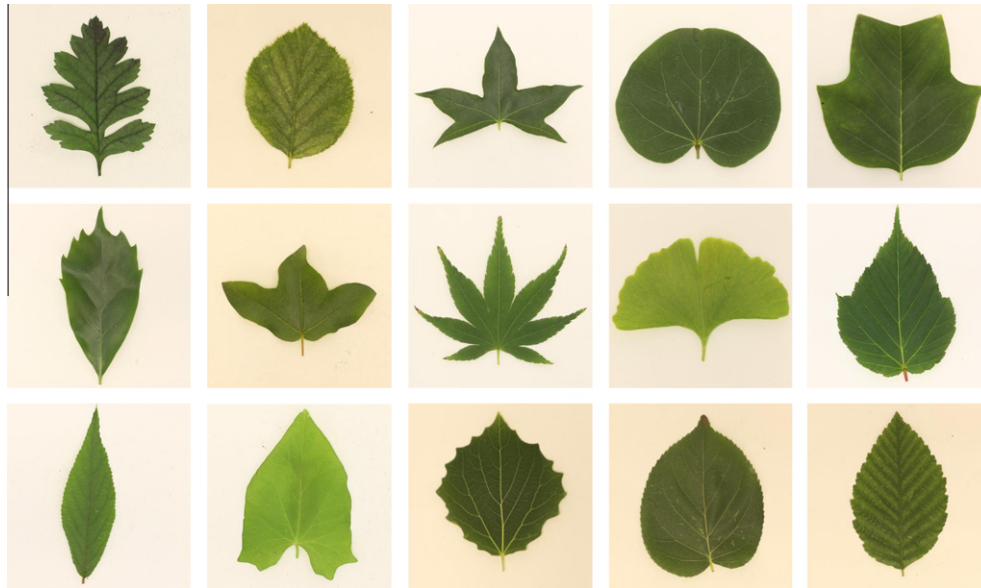


Fig. 3. Example of leaf shapes.

which has only four coefficients. This set of coefficients forms the Fourier descriptor, with higher numbers of harmonics providing more precise descriptions. (Hearn (2009) suggests that 10 Fourier harmonics are necessary to accurately represent leaf shape to distinguish between a range of species.) Typically, principal component analysis (PCA) is then applied to the descriptor, to reduce dimensionality and aid discrimination. An early example of this approach is by White, Prentice, and Verwijst (1988), who found EFA to be superior to landmark measures, chain codes and moment invariants when characterizing leaf outlines. Elliptic Fourier descriptors can easily be normalised to represent shapes independently of their orientation, size or location, easing comparison between shapes.

One advantage of EFDs is that a shape can be reconstructed from its descriptor, as shown in Fig. 4. A useful method for helping to explain shape variation is to reconstruct the shape for some “average” descriptor, and then to create reconstructions from this descriptor as it is modified along the first few principal components.

McLellan and Ender (1998) compared Fourier analysis with several other methods for describing leaf shape. They demonstrate

that Fourier analysis can discriminate successfully between various leaf groups. They also point out that few landmarks are readily identifiable on most leaves (see Section 2.1.3), except perhaps those that have regular lobes, making EFA methods suitable. They do note however that none of the methods they considered was greatly superior to any other.

Hearn (2009) used a combination of Fourier analysis and Procrustes analysis (Goodall, 1991) (a simple shape registration method, based on rotation, translation and scaling) to perform species identification using a large database of 2420 leaves from 151 different species. Du, Huang, Wang, and Gu (2005) successfully combined Fourier analysis with a radial basis function neural network to identify plant species from leaf images. Other recent examples of the use of EFDs to analyze leaf shape include Andrade, Mayo, Kirkup, and Van Den Berg (2008), Furuta, Ninomiya, Takahashi, Ohmori, and Ukai (1995), Neto, Meyer, Jones, and Samal (2006) and Lexer et al. (2009).

A closely related method is “eigenshape analysis”. Here, the sequence of angular deviations that define a contour is measured, typically being normalized by choosing a common starting point

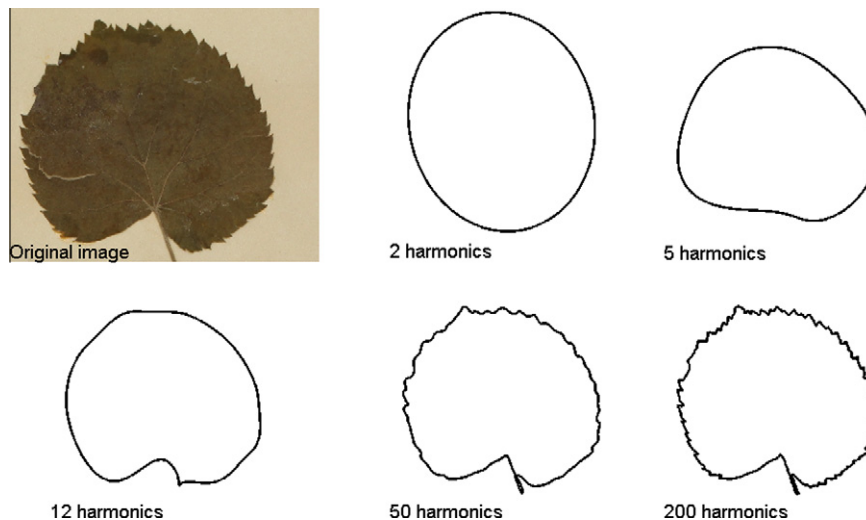


Fig. 4. An example of elliptic Fourier analysis. As more harmonics are used to reconstruct the original outline, more detail is preserved.

defined by a landmark. Singular value decomposition is then used to identify the principal components (MacLeod, 1999), which can be used as inputs to a subsequent classifier or for comparison. Ray has extended this work and applied it to leaf shape analysis (Ray, 1992). This work consisted of dividing the outline into several segments using recognizable landmarks (see Section 2.1.3), and then analyzing each segment using singular value decomposition. One difficulty with this approach is the problem of identifying homologous landmarks in leaves. While this can be difficult within a single species, it is often impossible between species, as we discuss further in Section 2.1.3.

2.1.2. Contour signatures

A number of methods make use of contour signatures. A contour signature for a shape is a sequence of values calculated at points taken around a leaf's outline, beginning at some start point, and tracing the outline in either a clockwise or anti-clockwise direction. One of the most straightforward of these is the centroid-contour distance (CCD). This signature consists of the sequence of distances between the centre of the shape, and the outline points. Other such signatures include the centroid-angle, and the sequence of tangents to the outline. As with EFDs, the aim of creating contour signatures is to represent the shape as a vector, independent of orientation and location. Normalisation can also be applied to also enforce independence of scale.

Meade and Parnell (2003) attempted to increase accuracy when applying the CCD to leaves by correlating the frequency of points for measurement with the extent of curvature, whilst Wang, Chi, Dagan, and Wang (2000), Wang, Chi, and Dagan (2003) applied a thinning-based method to the shape to identify consistent start points for the CCD, avoiding the need to align the signatures before they can be compared. Ye and Keogh (2009) used time-series shapelets. These are local features found in a contour signature that can be matched, rather than needing to compare entire signatures, and allows existing time-series analysis methods to be applied.

One major difficulty for boundary-based methods, to which contour signatures are particularly sensitive, is the problem of "self-intersection". This is where part of the leaf overlaps another part of the same leaf, and can result in errors when tracing the outline unless particular care is taken. Self-intersection occurs quite often with lobed leaves, and may, moreover, not even occur consistently within a particular species. One attempt to overcome this problem was made by Mokhtarian and Abbasi (2004). They assumed that darker areas of the leaf represented regions where overlap occurred, and used this to try to extract the true outline. They then used the curvature scale space method (CSS) to compare outlines. The main limitation of this method is that it will only work with thin and/or backlit leaves, where sufficient light can pass through the leaf to create the darker areas of overlap.

2.1.3. Landmarks and linear measurements

Another common morphometric method is the use of landmarks and linear measurements. A landmark is a biologically definable point on an organism, that can be sensibly compared between related organisms. It typically requires domain-specific expert knowledge to choose a suitable set of landmarks. In some cases, these are homologous points, but may instead be local maxima or minima of a boundary, as discussed by Bookstein (1986). Linear and angular measurements between them can then be used to characterize the organism's shape. Landmark methods have been successfully applied to various animal species, and have the advantage of being easy for a human to understand. "Traditional morphometrics" analyzes measurements such as the overall length and width of an object, in contrast to "geometric morphometrics", which uses either outlines (such as methods discussed in

Section 2.1.1) or specific landmarks and the distances between them (Adams, Rohlf, & Slice, 2004).

Haigh, Wilkin, and Rakotonasolo (2005) used leaflet lengths and widths along with measurements of flowers and petioles to differentiate two closely related species of *Dioscorea*. Jensen, Ciofani, and Miramontes (2002) studied three species of *Acer* using the angles and distances between the manually located lobe apices and sinus bases. Warp deformation grids were also used to study variation. Young, Dickinson, and Dengler (1995) used leaf landmarks to compare plants of a single species grown in different conditions. The plants were also imaged at different ages to discover when the method would have the best discriminatory ability. A related method is the inner-distance measure, a metric based on the lengths of the shortest routes between outline points without passing outside of the shape, which was used by Ling and Jacobs (2007).

A number of limitations exist, however, when applying landmark methods to leaves or other plant organs. The first of these is the difficulty of automatic extraction. For example the leaf's apex (tip) may be hard to distinguish from the tip of a lobe, whilst the appearance of the insertion point (where the petiole, or leaf stalk, meets the leaf blade) may vary greatly depending on the base angle and how the petiole has been cut during specimen preparation. Furthermore, even the length of a leaf may be hard to measure if the leaf is asymmetrical and the main vein does not align with the shape's primary axis. For these reasons, studies involving landmarks and linear measurements (including those mentioned above) have often involved manual data extraction by experts, severely limiting the scale of any system based on them.

The other major problem here is the inconsistency in available landmarks between different species or other taxa, as demonstrated by the variety of shapes shown in Fig. 3. Indeed, the only landmarks present in almost all leaves are the apex and the insertion point (Fig. 2), and in the case of peltate leaves (where the stalk is connected near the middle of the blade), the latter does not even appear in the outline shape. As a result, most of the studies using landmarks concentrate on specific taxa where the required features are known to be present. One exception is work by Corney, Clark, Tang, and Wilkin (2012) who describe software that automatically analyses herbarium images to identify the leaf tip and petiole insertion point landmarks. From these, and from the whole leaf boundary, it automatically extracts features including the leaf blade's length and width, as well as area, perimeter and various other shape features.

One of the most significant developments in comparative biology in the last 30 years has been the development of phylogenetic reconstruction methods using morphological data, and latterly nucleic acid or protein sequence data. These methods differ from those dealt with elsewhere in this paper in that they use only shared derived characters to infer (phylogenetic) relationship rather than using total overall resemblance for identification or species delimitation (compare cladistics *versus* phenetics; see Table 1). The concept of homology has particular importance in cladistics and is perhaps more tightly defined (Patterson, 1982). There has been theoretical debate over the use of continuous and hence morphometric morphological character data in cladistics. Several approaches have been suggested, such as that of Thiele (1993). Zelditch, Swiderski, and Fink (2000) even attempted to use geometric morphometric methods, such as partial warps, to acquire novel phylogenetic character data in fish, although such techniques have not been widely used in systematics as a whole or been taken up by plant systematists.

2.1.4. Shape features

Similar to linear measurements are shape features, which are also typically limited to analysing the outline of a shape. These are various quantitative shape descriptors that are typically

intuitive, easy to calculate, and applicable to a wide variety of different shapes. Commonly used features include the shape's aspect ratio, measures of rectangularity and circularity, and the perimeter to area ratio, amongst others. Some studies have also used more leaf-specific features, for example Pauwels, de Zeeuw, and Rangelova (2009) uses a measure of "lobedness". A more general set of features are "invariant moments", which are statistical descriptors of a shape that are invariant to translation, rotation and scale (Hu, 1962, Teague, 1980). Flusser, Suk, and Zitov (2009) provides a modern, comprehensive review of moment invariants for pattern recognition generally.

When analyzing leaves, Lee and Chen (2006) argue that "region-based features", such as compactness and the aspect ratio, are more useful than outline contour features because of the difficulty in identifying meaningful landmark points, or in registering different contours against each other. They found that a simple nearest-neighbour classifier using region-based features produced better results than a contour based method, at least on the 60 species they used as a test case.

Once such a set of features has been extracted from the images, a variety of classifiers can be used in their analysis. A "move median centres" (MMC) hypersphere classifier was developed by Du and colleagues (Du, Wang, & Zhang, 2007, Wang et al., 2008) that uses a series of hyperspheres to identify species in a space defined by a set of shape features and invariant moments. Another study using shape features was carried out by Wu et al. (2007), who used an artificial neural network to identify 32 species of Chinese plants from images of single leaves, and compared the results against a number of other classifiers.

While shape features have achieved some positive results, they are of limited use for aiding understanding of variation. Although the effects of some features may be obvious, such as changes to the height-to-width ratio, variation in other features may be hard to understand because of the difficulty or impossibility of reconstructing shapes from features. For example, the perimeter to area ratio provides a measure of the "complexity" of a shape, but there are many ways in which a leaf might be altered to produce the same change in this value, without affecting the values of many other common shape features.

A more general risk with shape features is that any attempt to describe the shape of a leaf using only (say) 5–10 features may oversimplify matters to the extent that meaningful analysis becomes impossible, even if it is sufficient to assign a small set of test images to the correct categories. Furthermore, many such single-value descriptors are highly correlated with each other (McLellan & Endler, 1998), making the task of choosing sufficient *independent* variables to distinguish categories of interest especially difficult.

2.1.5. Polygon fitting and fractal dimensions

The fractal dimension of an object is a real number used to represent how completely a shape fills the dimensional space to which it belongs. This can provide a useful measure of the "complexity" of a shape, which may then be used as an input feature for a classifier, for example. There are many ways to define and calculate an object's fractal dimension, with the Minkowski-Bouligand method being a popular choice due to its precision and the existence of a multi-scale version.

A few attempts have been made to use fractal dimensions to identify leaves. McLellan and Endler (1998) used the fractal dimension as a single value descriptor alongside other descriptors. Plotze et al. (2005) used the positions of feature points in the curves produced by the multi-scale Minkowski-Bouligand fractal dimension, whilst Backes and Bruno (2009) also used the multi-scale Minkowski-Bouligand method, but compared Fourier descriptors calculated for the curves. Bruno, de Oliveira Plotze, Falvo, and de Castro (2008) compare box-counting and multi-scale Minkowski

estimates of fractal dimension, and used linear discriminant analysis to identify a number of plant species. McLellan and Endler (1998) showed that fractal dimension tends to be highly correlated with the perimeter to area ratio (or "dissection index"), suggesting it is of limited additional benefit.

As with shape features, whilst some good results have been achieved, with Plotze et al. (2005) reporting a 100% identification rate on a small database of 10 species of *Passiflora*, their usefulness in explaining variation is somewhat limited. Given the wide variety of leaf shapes present (e.g. Fig. 3), characterizing shape by any single measure of complexity may discard too much useful information, suggesting that fractal dimension measures may only be useful in combination with other features.

Du, Huang, Wang, and Gu (2006) created polygonal representations of leaves, and used these to perform comparisons, while Im, Nishida, and Kunii (1999) represented leaf outlines as a series of super-imposed triangles, which could then be normalized and registered against each other for comparison. The method was shown to correctly identify 14 Japanese plant species, but relies on a number of heuristic assumptions, which may limit the method's applicability to a more general task.

2.2. Venation extraction and analysis

After their shape, the next most studied aspect of leaves is the vein structure, also referred to as the venation. Veins provide leaves with structure and a transport mechanism for water, minerals, sugars and other substances. The pattern of veins in a leaf can be used to help identify a plant. Although the fine detail may vary, the overall pattern of veins is conserved within many species. Veins are often clearly visible with a high contrast compared to the rest of the leaf blade. See Figs. 5 and 6.

A wide variety of methods have been applied to the extraction of the vein networks, although arguably with limited success thus far. Clarke et al. (2006) compared a couple of simple methods (a scale-space analysis algorithm and a smoothing and edge detection algorithm) to results achieved manually using Adobe Photoshop. They report the quality of the results as judged by some expert botanists, and although the manual results were preferred, the results showed some hope for automatic methods, for at least some species.

Cope, Remagnino, Barman, and Wilkin (2010a) used genetic algorithms to evolve classifiers for identifying vein pixels. These were robust and capable of recognizing primary and secondary venation to a high degree of accuracy. Li, Chi, and Feng (2006) successfully extracted the venation from leaf sub-images using Independent Component Analysis (ICA) (Comon, 1994), though when used on whole leaves, the results were no better than using a simple Prewitt edge detection operator. Artificial ant swarms were used by Mullen, Monekosso, Barman, Remagnino, and Wilkin (2008) to trace venation and outlines in leaves via an edge

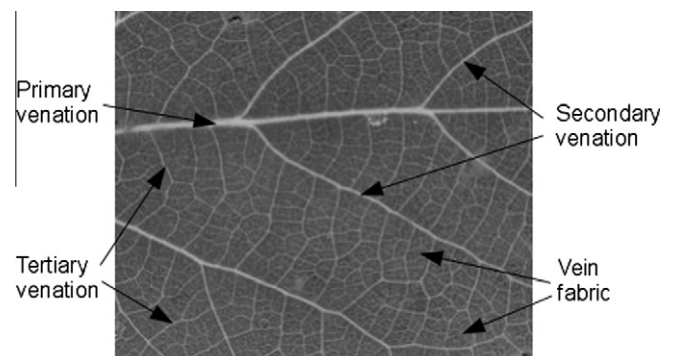


Fig. 5. Example of leaf vein structure.



Fig. 6. Examples of variety of vein structures.

detection method. Some of the best vein extraction results were achieved by Fu and Chi (2006) using a combined thresholding and neural network approach. Their experiments were, however, performed using leaves which had been photographed using a fluorescent light bank to enhance the venation, and such images are not generally available. Kirchgessner, Scharr, and Schurr (2002) used a vein tracing method with extracted veins represented using b-splines, whilst Plotze et al. (2005) used a Fourier high-pass filter followed by a morphological Laplacian operator to extract venation.

Whilst there have been several attempts at extracting venation, there have been fewer attempts to analyze or compare it, with most of these using synthetic or manually extracted vein images. Park, Hwang, and Nam (2008) used the pattern of end points and branch points to classify each vein structure as one of the main venation types (see Fig. 6), and Nam, Hwang, and Kim (2008) performed classification on graph representations of veins. Further evaluation is required before the general value of venation analysis can be determined.

2.3. Leaf margin analysis

The leaf margin, the outer edge of the lamina, often contains a pattern of “teeth” – small serrated portions of leaf, distinct from the typically larger and smoother lobes (see Fig. 7 for examples). Despite being a useful feature of leaves for botanists to use when describing leaves, the margin has seen very little use in automated leaf analysis. Indeed, it has been claimed that “no computer algorithm can reliably detect leaf teeth” (Royer, Wilf, Janesko, Kowalski, & Dilcher, 2005) as yet. This may be due to the fact that teeth are not present in all species of plant; that they are damaged or missing before or after specimen collection; or due to the difficulty in acquiring quantitative measurements automatically. Nonetheless, teeth are an important feature of many plant species, with botanists using qualitative descriptors of the tooth curvature (Ellis et al., 2009). The size and number of teeth can also be useful

indicators of climate and of growth patterns (Royer & Wilf, 2006), and are even used to make inferences about prehistoric climates using fossilized leaves (Ellis et al., 2009).

Studies using the leaf margin normally combine it with other features and measurements. Clark (2004), Clark (2007), Clark (2009) and Rumpunen and Bartish (2002) both use manually taken measurements such as tooth length and width (“pitch”), used alongside various linear shape measurements. Clark (2004) showed that a multi-layer perceptron outperformed a computer-generated taxonomic key for identifying species from morphological traits. Clark (2009) used a self-organizing map to identify species boundaries from similar morphological traits. McLellan and Endler (1998) used the sum of the angles between lines connecting adjacent contour points along with other single value leaf features, and Wang et al. (2003) compared histograms of the angles at points spread around the contour.

For taxa that possess teeth, if sufficient undamaged leaves are available, then the area of the toothed margin region and the size and numbers of teeth may be useful characters to measure. One possibility for future work is to combine vein analysis (Section 2.2) with margin analysis, as teeth often have small veins running to their tips. Margins that are broken or insect-damaged may look superficially like teeth, but are less likely to have the same patterns of veins. Clearly, for taxa that do not possess teeth, other methods must be used – as noted in Section 1.1, different analytical tasks may require different features.

2.4. Leaf texture analysis

Besides analysing outlines, a number of both traditional and novel texture analysis techniques have been applied to leaves. Backes et al. have applied multi-scale fractal dimensions (Backes & Bruno, 2009) and deterministic tourist walks (Backes, Gonçalves, Martinez, & Bruno, 2010) to plant species identification by leaf texture, although their experiments involved very limited datasets which makes them hard to evaluate. Casanova, de Mesquita Sá Junior,



Fig. 7. Examples of variety of leaf margins.

and Bruno (2009) used an array of Gabor filters on a larger dataset, calculating the energy for the response of each filter applied, and achieved reasonable results, whilst Liu et al. have presented a method based on wavelet transforms and support vector machines (Liu, Zhang, & Deng, 2009). Cope, Remagnino, Barman, and Wilkin (2010b) achieved an 85% identification rate on 32 species of *Quercus* using the co-occurrences of different scale Gabor filters. Other techniques used include Fourier descriptors and grey-scale co-occurrence matrices.

Whilst the above studies were all performed on texture windows acquired using traditional imaging techniques (i.e. cameras and scanners), Ramos and Fernandez (2009) used images acquired using a scanning electron microscope (SEM), and Backes, de Mesquita Sá Junior, Kolb, and Bruno (2009) used magnified cross-sections of the leaf surface epidermis (the outermost layer of cells). While these provide interesting results, such images are not available on a large scale.

Where texture is preserved in a specimen, such analysis may prove useful, especially when combined with outline-based shape analysis.

2.5. Other lamina-based methods

There have been a few other studies which have used the leaf lamina (surface), or features present on it, in ways different from those already discussed. Gu, Du, and Wang (2005) processed the laminae using a series of wavelet transforms and Gaussian interpolation to produce a leaf “skeleton”, which is used to calculate a number of run-length features: measure of short runs; measure of long runs; distribution of grey-scales; distribution of lengths and the percentage of runs.

Qualitative leaf hair descriptors were used by Clark (2009) as one feature in a self-organizing map. These were manually identified and described, and pose a problem for automated systems due to their three-dimensional nature which makes positive identification from a two-dimensional image difficult. Surface glands are another potentially useful lamina feature that have been largely ignored thus far in computational methods as far as we are aware.

One intriguing option is to apply 3D imaging and modelling methods to leaf shapes (or to flowers; see below). Ma, Zha, Liu, Zhang, and Xiang (2008) describe one such method which uses volumetric information from a 3D scanner to reconstruct leaves and branches of plants, though it is not clear how this would work on a large scale system. Teng, Kuo, Chen, and segmentation (2009) combine several 2D photos of the same scene to extract 3D structure, and use the 2D and 3D information together to segment the image, using normalized cuts, finding the leaf boundary. They then use centroid contour distance (CCD, as discussed in Section 2.1.2) to classify leaves into broad classes, such as palmate or cordate (see Fig. 3). Similar work is described by Song, Wilson, Edmondson, and Parsons (2007), where stereo image pairs were analyzed using stereo matching and a self-organizing map. The resulting surface models contained sufficient detail to allow measurements of leaf and flower height, as well as shape.

The laminae of most leaves contain many stomata, which are pores that open or close to control gaseous exchange including water loss. It has been shown that the size and distribution of these is closely related to the climate and to CO₂ concentrations in particular. Royer (2001) reviews a wide range of data that shows that the density of stomata on fossil leaves is inversely related to local CO₂ concentrations over long timescales. Hetherington and Woodward (2003) discuss this and the effect of changing environmental factors over shorter timescales, as well as discussing the morphology of different types of stomata. Zarinkamar (2007) presents a thorough botanical description and corresponding (manual) measurements of a wide range of stomata shapes found across more

than 300 species, and argues that such measurements can be used to aid taxonomic classification, as well as to monitor changes in the local environment. Fernandez (2008) presents a method to mathematically analyse digital microscope images of leaf stomata. They use various measures of correlation and entropy to characterize the patterns and textures found, and use PCA to help visualize and group the results. We are not aware of attempts to perform automatic species identification or related tasks based on image processing of stomata. However, it seems to us that by combining the separate pieces of work discussed here, it should be possible to do so.

2.6. Flowers and other plant organs

Although the focus of this paper, leaves are not the only plant organs on which image processing and morphometric techniques have been applied. Traditional keys often make use of descriptions of flowers and/or of fruits, but these are often only available for a few days or weeks of the year.

A number of methods have been proposed to identify plants from digital images of their flowers. Although colour is a more common distinguishing feature here, many methods used to analyze leaf shape can also be used (see earlier sections). Nilsback and Zisserman (2007) combined a generic shape model of petals and flowers with a colour-based segmentation algorithm. The end result was a good segmentation of the image, with species identification left to future work.

Das, Manmatha, and Riseman (1999) demonstrated the use of colour alone to identify a range of flowers in a database related to patents covering novel flower hybrids. Their method allows the database to be searched by colour name or by example image, although no shape information is extracted or used. A colour-histogram segmentation method was used by Hong, Chen, Li, Chi, and Zhang (2004) and then used with the centroid contour distance (CCD; see Section 2.1.2) and angle code histograms to form a classifier. They demonstrated that this method works better than using colour information alone to identify a set of 14 species. This again suggests that outline shape is an important character to consider, especially in combination with other features.

Elliptical Fourier descriptors (Section 2.1.1) were used by Yoshioka, Iwata, Ohsawa, and Ninomiya (2004) to study the shape of the petals of *Primula sieboldii*, whilst Wilkin (1999) used linear measurements of floral organs, seeds and fruits as well as leaves and PCA methods to investigate whether a closely related group of species in Africa were morphologically distinct or not. He discovered that they in fact formed a single morphological entity and hence all belonged to one species. Gage and Wilkin (2008) used EFA on the outlines of tepals (elements of the outer part of a flower, such as petals and sepals) of three closely related species of *Sternbergia* to investigate whether they really formed distinct morphological entities. Clark (2009) used linear measurements of bracts, specialized leaf-like organs, in a study of *Tilia* using self-organizing maps, and Huang, De-Shuang, Du, Quan, and Guo (2006) analyzed bark texture using Gabor filters and radial basis probabilistic neural networks.

At a smaller scale, the growth of individual grains of barley has been modelled by 3D reconstruction from multiple 2D microscopic images (Gubatz, Dercksen, Brüß, Weschke, & Wobus, 2007). This allowed both “virtual dissecting” of the grains as an educational tool, and also visualization of gene expressions via mRNA localization. At a smaller scale still, Oakley and Falcon-Lang used a scanning electron microscope (SEM) to analyze the vessels found in fossilized wood tissue (Oakley & Falcon-Lang, 2009). They used principal component analysis (PCA) to identify two distinct “morphotypes”, which correspond to one known and one novel species of plant growing in Europe around 95 million years ago.

Moving underground, a number of studies have used image processing techniques to analyze root structures in the “rhizosphere” (the region that roots grow in, including the soil, soil microbes, and the roots themselves). For example, Huang, Jain, Stockman, and Smucker (1992) used digital images of roots captured by placing a small camera inside a transparent tube placed beneath growing plants. They then used expert knowledge of root shapes and structures (such as roots being elongated and having symmetric edges), to combine multiple sources of information and to fit polynomial curves to the roots, and use a graph theoretic model to describe them. More recently, Zeng, Birchfield, and Wells (2008) used image intensity to distinguish root pixels from soil pixels. They then used a point process to combine and connect segments to efficiently identify complete root systems.

These studies show that while the clear majority of botanical morphometrics research has focused on leaves, due to their ready availability and use for discriminating between taxa, other plant organs, when available, should not be ignored.

3. Systems for species identification, robotic agriculture and botany

In this section, we move beyond discussing specific algorithms in isolation and methods designed for the laboratory, and consider a number of complete systems and prototypes, designed for practical use in the field. In order to have an impact in the real world, it is important to demonstrate that algorithm such as those described earlier can be applied in practice, and can scale up from a few idealized examples to larger and more complex problems. We review systems designed to identify species from plant images; several agricultural applications; and scientific research tools regarding species variation and distribution, and how this relates to the climate.

3.1. General purpose species identification

Plant identification is currently particularly important because of concerns about climate change and the resultant changes in geographic distribution and abundance of species. Development of new crops often depends on the incorporation of genes from wild relatives of existing crops, so it is important to keep track of the distribution of all plant taxa. Automated identification of plant species, for example using leaf images, is a worthwhile goal because of the current combination of rapidly dwindling biodiversity, and the dearth of suitably qualified taxonomists, particularly in the parts of the world which currently have the greatest numbers of species, and those with the largest number of “endemics” (species restricted to that geographic area).

The species to which an organism belongs is often regarded as its most significant taxonomic rank. Accurately identifying an organism to species level allows access to the existing knowledge bases that are linked to that specific name, such as what other species the taxon in question may breed or hybridize with, what it uses are, and so on.

A number of systems have been developed that aim to recognize plant species from the shapes of their leaves. One such plant identification system is described by Du et al. (2006). They argue that any *global* shape-based method is likely to perform poorly on images of damaged or overlapping leaves because parts of the leaf perimeter are missing or obscured. Instead, they suggest that local shape-based methods are more robust for this type of task. Their system matches leaves from images by fitting polygons to the contour and using a modified Fourier descriptor with dynamic programming to perform the matching. It aims to be robust with regard to damaged or overlapping leaves, as well as blurred or

noisy images. They claim a 92% accuracy for their method on one sample of over 2000 “clean” images, representing 25 different species, compared with 75%–92% for other methods, and that their method is more robust for images of incomplete or blurred leaves.

The increasing power and availability of cheap hand-held computers, including personal digital assistants (PDAs) and smart phones, has led to a number of prototype applications. The goal of allowing users, both professional botanists and interested amateurs, to go out into the field and identify plant species using an automated system is a highly desirable goal, although the task is challenging, not least because of the very large number of plant species that may be encountered.

One major and ongoing project aims to produce an “electronic field guide” to plants in the USA (Agarwal et al., 2006). The user can take a photograph of a single leaf on a plain background, and the system will display images of twenty plant species that have the closest match. Shape similarity is measured using the Inner-Distance Shape Context algorithm, which extends the shape context work of Belongie, Malik, and Puzicha (2002). A related prototype from the same project includes an “augmented reality” feature (Belhumeur et al., 2008), and provides a visual display of a herbarium specimen for side-by-side comparison to the plant in question (White, Feiner, & Kopylec, 2006). Both systems are currently limited to certain geographical regions of the USA, although there are plans to extend its functionality to cover the rest of the country, and eventually further afield.

The CLOVER system (Nam, Hwang, & Kim, 2005) allows users to provide a sketch or a photograph of a leaf using a hand-held computer, which then accesses a remote server. The server retrieves possible matches based on leaf shape, using several shape matching methods including an enhanced version of the minimum perimeter polygons algorithm, and returns the matches to the device to display to the user. The information on the server included data extracted from over 1000 images from a Korean flora, itself created by an expert botanist. The prototype described has been demonstrated to work effectively at recognizing plants from leaves, with the inevitable trade-off between recall and precision.

A similar system uses fuzzy logic and the centroid-contour distance to identify plant species from Taiwan (Cheng, Jhou, & Liou, 2007). However, this requires the user to select various characteristics of the plant from a series of menu options, rather than using morphometric analysis directly.

Each of these general purpose prototypes has been demonstrated to work successfully on at least a small number of species, and under more or less stringent conditions. Currently, we know of no such system that is available for everyday use, although interest remains high (Lipske, 2008).

3.2. Agriculture

Rather than trying to identify a plant as belonging to one particular species, it is sometimes sufficient to recognize a plant as “good” or “bad”, without needing to be concerned about the exact taxon to which it belongs. One goal of automated or “precision” agriculture (Burgos-Artizzu, Ribeiro, Tellaache, Pajares, & Fernández-Quintanilla, 2010) is to allow targeted administration of weed killer, fertilizer or water as appropriate from an autonomous robotic tractor, not least to minimize the negative impact on the environment of large scale agriculture. To do this, the system must obviously identify plants as belonging to one category or the other, such as “weed” vs. “crop”.

As is often the case with machine vision systems, variable lighting conditions can make image processing very hard. One proposed solution is to control the lighting by building a light-proof “tent” that can be carried on wheels behind a tractor, and which contains lamps inside it along with a camera. One such system successfully

distinguishes between crop plants (cabbages and carrots) and weed plants (anything else) growing in field conditions (Hemming & Rath, 2001). Whether carrying round such a bulky tent is feasible or not on a larger scale, it is certainly not ideal.

A similar system uses rails to guide a vehicle carrying the camera along carefully laid out plots (Gebhardt, Schellberg, Lock, & K+hbauch, 2006). Rather than carrying its own lights, the system is only used under standardized illumination conditions (e.g. bright but overcast). This system extracts shape features such as leaf circularity and area and uses a maximum likelihood estimator to identify leaves that are weeds (specifically dock leaves, *Rumex obtusifolius*) in grassland, with around 85%–90% accuracy. A different system to identify dock leaves is described by Seatović (2008), which uses a scanning laser mounted on a wheeled vehicle to generate 3D point clouds. These are then segmented to separate out leaves from their background, and a few simple rules, based on leaf size, are used to distinguish the dock leaves from other leaves in the meadow.

A related attempt to distinguish weeds, crops and soil in field conditions uses morphological image processing (Soille, 2000). This attempts to identify the centre of each leaf by using colour threshold segmentation and locating the leaf veins. The system locates the veins using a combination of morphological opening and hierarchical clustering. The final classification makes use of *a priori* knowledge about features of the target plant species, such as known leaf sizes and venation patterns. A similar system combines morphological processing with an artificial neural network classifier has also been suggested (Pan & He, 2008). This used a radial basis function network to distinguish grass and other weeds from crops. A combination of colour segmentation and morphological programming has also been used towards the development of a robotic cucumber harvester (Qi, Yang, Bao, Xun, & Zhang, 2009).

A variety of methods to distinguish various crops from weeds and soil are discussed by Burgos-Artizzu et al. (2010), including colour segmentation and morphological processing, and the use of genetic algorithms to optimise these methods. The paper also provides a useful overview of research into “precision agriculture”, which aims to use modern technology to optimize crop production, allowing for local variation in soil, landscape, nutrients and so on.

3.3. Intraspecific variation, geographical distribution, and climate

It has long been known that the climate in which a plant grows has an effect on the shape of its leaves (Royer et al., 2005). Recent work has extended this by using digital image analysis to enhance the botanical and climatic measurements, which can then be analysed using existing knowledge discovery techniques. Huff, Wilf, and Azumah (2003) collected leaves from temperate and tropical woodlands. They analyzed the leaves and measured the shape factor, and found a correlation with the mean annual temperature. The work was then extended to a wider variety of environments (17 in total) in North America (Royer et al., 2005). Here, a variety of simple digital image analysis methods were used to semi-automatically measure features such as leaf blade area, tooth area, number of teeth, and major and minor axis lengths. These features were then compared to climatic measurements from the different field locations. Finally, correlations between leaf shape and climate were measured. They confirmed previous findings that plants growing in colder environments tend to have more teeth and larger tooth areas than similar plants growing in warmer environments. One of the goals of this body of work is to support analysis of leaf fossils, with the aim of estimating paleoclimatic conditions. By establishing how leaves from living plants have shapes that correlate with their environments, it is hoped that fossil leaf shapes can indicate how the Earth’s climate has changed in the past, at both a global and a local scale.

In botany, identifying taxon boundaries is often as important as identifying to which taxa a particular specimen belongs. An early study by Dickinson, Parker, and Strauss (1987) used manual digitization (via a tablet) to identify landmarks on cross-sections of leaves, and principal component analysis to analyze the data. They identified both geographic variation between collection sites and also identified intermediate forms of specimens, suggesting various hybridizations had occurred. As mentioned earlier, work by Wilkin (1999), Gage and Wilkin (2008) used morphometric analysis to identify species boundaries.

4. Conclusions

In this paper we have discussed a number of species identification systems that rely on both domain knowledge and on a wide range of morphometric methods. It should be clear that no single method provides a panacea for all problems, but rather that appropriate methods must be chosen for each task at hand. Plants are extremely diverse in shape, size and colour. A method that works very well on one group may rely on features that are absent in another taxon. For example, landmarks may be readily definable and identifiable for some taxa, such as those with distinctive lobes, but not for others.

Given the large scale nature of botanical morphometrics and image processing, automation is essential. Any system that requires significant manual effort, for example in tracing leaf outlines, is unlikely to be practical when scaled up to thousands of specimens. Despite this, in some cases the user may remain involved in the process with no great cost: if an electronic field-guide provides say ten predictions of species, rather than one, the user may be able to readily choose the most likely answer (Agarwal et al., 2006). Related to this is the issue of processing speed. The user of a hand-held field-guide may require responses interactively and so (near) instantaneously, whereas if a tool is to be used on a large set of images in a botanical laboratory, it may be acceptable to wait overnight for a comprehensive result – assuming no human interaction is needed.

A complement to plant identification is plant modelling. Various algorithms have been developed that reproduce the typical branching structure of plants, especially during growth, such as L-systems (Lindenmayer systems) (Prusinkiewicz & Lindenmayer, 1990). These can be extended to incorporate the effect of different growing conditions, environmental impact and the genetic origins of plant forms (Prusinkiewicz, 2004). Such models typically have relatively few parameters, and by varying these, a wide variety of virtual plants can be generated. If such generation can be matched to data derived from biological plants, then it may be possible to model the process by which real, observed plants have been produced, which in turn could be of great interest to plant science, not least for species identification. As far as we are aware, such work has not been undertaken, possibly due to the computational complexity of the matching process.

We finally return briefly to some of the challenges presented in Section 1.1. When presented with a very large number of classes, and classes that are often distinguished by using distinct sets of features, we suggest two broad solutions. First, one could restrict the task to consider only a small number of classes – e.g. developing a system to identify only two or three species rather than hundreds or thousands. A number of papers discussed earlier reflect this, deliberately or inadvertently. Second, one could develop a hierarchical system, possibly following taxonomic models of evolutionary inheritance, and consider *at each stage* only a small number of classes and a restricted set of features. Such a model could be developed in a gradual, modular fashion, and could be achieved collaboratively. Of course, the former solution may well

be a component of the latter. We also note that interdisciplinary work can be fraught with communication problems unless great care is taken to ensure that terminology is used consistently. We hope that this paper goes some way to reducing this risk.

The use of computational, morphometric and image processing methods to analyse leaf images is particularly timely. E. O. Wilson has argued for the creation of an “encyclopedia of life” (Wilson, 2003) – a web page for every species of life on Earth – and with it, the embrace of new technologies, such as digital images of plant specimens provided by the world’s herbaria. We believe that automated image processing and morphometrics can help to fulfil this goal both by helping to define species more effectively and providing rapid, web-based species identifications.

Acknowledgements

This work is kindly funded by the Leverhulme Trust (grant F/00 242/H), as part of MORPHIDAS: the Morphological Herbarium Image Data Analysis research project. We thank Sarah Barman, Lilian Tang, Ken Bailey and Richard White for their helpful comments. Unattributed photographs were taken by James Cope.

References

- Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: Ten years of progress following the “revolution”. *Italian Journal of Zoology*, 71, 5–16.
- Agarwal, G., Belhumeur, P., Feiner, S., Jacobs, D., Kress, W. J., Ramamoorthi, R., et al. (2006). First steps toward an electronic field guide for plants. *Taxon*, 55, 597.
- Andrade, M., Mayo, S. J., Kirkup, D., & Van Den Berg, C. (2008). Comparative morphology of population of *Monstera Adans.* (Araceae) from natural forest fragments in northeast Brazil using elliptic Fourier analysis of leaf outlines. *Kew Bulletin*, 63, 193–211.
- Backes, A. R., & Bruno, O. M. (2009). Plant leaf identification using multi-scale fractal dimension. In *International Conference On Image Analysis And Processing* (pp. 143–150). Berlin/ Heidelberg: Springer.
- Backes, A. R., de Mesquita Sá Junior, J. J., Kolb, R. M., & Bruno, O. M. (2009). Plant species identification using multi-scale fractal dimension applied to images of adaxial surface epidermis. *Computer Analysis Of Images And Patterns*, 5702, 680–688.
- Backes, A. R., Gonçalves, W. N., Martinez, A. S., & Bruno, O. M. (2010). Texture analysis and classification using deterministic tourist walk. *Pattern Recognition*, 43, 685–694.
- Belhumeur, P., Chen, D., Feiner, S., Jacobs, D., Kress, W., Ling, H., Lopez, I., Ramamoorthi, R., Sheorey, S., White, S., & Zhang, L. (2008). Searching the world’s herbaria: A system for visual identification of plant species. *Computer Vision – ECCV 2008* (pp. 116–129). Springer.
- Belongie, S., Malik, J., & Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 509–522.
- Bishop, C. M. (2007). *Pattern Recognition and Machine Learning* (first ed.). Springer.
- Bookstein, F. L. (1986). Size and shape spaces for landmark data in two dimensions. *Statistical Science*, 1, 181–222.
- Bruno, O. M., de Oliveira Plotze, R., Falvo, M., & de Castro, M. (2008). Fractal dimension applied to plant identification. *Information Sciences*, 178, 2722–2733.
- Burgos-Artizux, X. P., Ribeiro, A., Tellaecha, A., Pajares, G., & Fernández-Quintanilla, C. (2010). Analysis of natural images processing for the extraction of agricultural elements. *Image and Vision Computing*, 28, 138–149.
- Carvalho, M. R., Bockmann, F. A., Amorim, D. S., Brando, C. R. F., Vivo, M., Figueiredo, J. L., et al. (2007). Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the Cybertaxonomic-Automation paradigm. *Evolutionary Biology*, 34, 140–143.
- Casanova, D., de Mesquita Sá Junior, J. J., & Bruno, O. M. (2009). Plant leaf identification using Gabor wavelets. *International Journal of Imaging Systems and Technology*, 19, 236–243.
- Cheng, S. C., Jhou, J. J., & Liou, B. H. (2007). PDA plant search system based on the characteristics of leaves using fuzzy function. In *New Trends in Applied Artificial Intelligence. LNAI* (vol. 4570, pp. 834–844). Springer.
- Clark, J. Y. (2004). Identification of botanical specimens using artificial neural networks. In *Proceedings of the 2004 IEEE symposium on computational intelligence in bioinformatics and computational biology, CIBCB’04* (pp. 87–94).
- Clark, J. Y. (2007). Plant identification from characters and measurements using artificial neural networks. In *MacLeod, N., (Ed.), Automated taxon identification in systematics: Theory, approaches and applications, CRC* (pp. 207–224).
- Clark, J. Y. (2009). Neural networks and cluster analysis for unsupervised classification of cultivated species of *Tilia* (Malvaceae). *Botanical Journal of the Linnean Society*, 159, 300–314.
- Clarke, J., Barman, S., Remagnino, P., Bailey, K., Kirkup, D., Mayo, S., et al. (2006). Venation pattern analysis of leaf images. *Lecture Notes In Computer Science*, 4292, 427–436.
- Comon, P. (1994). Independent component analysis: A new concept? *Signal Processing*, 36, 287–314.
- Cope, J. S., Remagnino, P., Barman, S., & Wilkin, P. (2010a). The extraction of venation from leaf images by evolved vein classifiers and ant colony algorithms. In *Advanced concepts for intelligent vision systems, Sydney, Australia, Springer-Verlag*.
- Cope, J. S., Remagnino, P., Barman, S., & Wilkin, P. (2010b). Plant texture classification using Gabor co-occurrences. In *International symposium on visual computing, Las Vegas, Nevada, USA, Springer-Verlag*.
- Corney, D. P. A., Clark, J. Y., Tang, H. L., & Wilkin, P. (2012). Automatic extraction of leaf characters from herbarium specimens. *Taxon*, 61, 231–244.
- Das, M., Manmatha, R., & Riseman, E. M. (1999). Indexing flower patent images using domain knowledge. *IEEE Intelligent Systems*, 14, 24–33.
- Dickinson, T. A., Parker, W. H., & Strauss, R. E. (1987). Another approach to leaf shape comparisons. *Taxon*, 1–20.
- Du, J., Huang, D., Wang, X., & Gu, X. (2005). Shape recognition based on radial basis probabilistic neural network and application to plant species identification. In *International Symposium On Neural Networks* (pp. 281–285). Berlin/ Heidelberg: Springer.
- Du, J.-X., Huang, D.-S., Wang, X.-F., & Gu, X. (2006). Computer-aided plant species identification (CAPSI) based on leaf shape matching technique. *Transactions Institute Of Measurement And Control*, 28, 275–284.
- Du, J.-X., Wang, X.-F., & Zhang, G.-J. (2007). Leaf shape based plant species recognition. *Applied Mathematics and Computation*, 185, 883–893.
- Ellis, B., Daly, D. C., Hickey, L. J., Johnson, K. R., Mitchell, J. D., Wilf, P., et al. (2009). *Manual Of Leaf Architecture*. Cornell University Press.
- Fernandez, J. M. (2008). Classification of textures in microphotographies of leaves epidermis. In *Proceedings of the national conference on undergraduate research (NCUR), Salisbury University, Maryland, USA*.
- Flusser, J., Suk, T., & Zitov, B. (2009). *Moments and Moment Invariants in Pattern Recognition*. John Wiley and Sons.
- Fu, H., & Chi, Z. (2006). Combined thresholding and neural network approach for vein pattern extraction from leaf images. In *IEE Proceedings of vision image and signal processing, Institution of Electrical Engineers* (Vol. 153, pp. 881–892).
- Furuta, N., Ninomiya, S., Takahashi, N., Ohmori, H., & Ukai, Y. (1995). Quantitative evaluation of soybean leaflet shape by principal component scores based on elliptic Fourier descriptor. *Breeding Science*, 45, 315–320.
- Gage, E., & Wilkin, P. (2008). A morphometric study of species delimitation in *Sternbergia lutea* (Alliaceae, Amaryllidoideae) and its allies *S. sicula* and *S. greuteriana*. *Botanical Journal of the Linnean Society*, 158, 460–469.
- Gebhardt, S., Schellberg, J., Lock, R., & K+hbauch, W. (2006). Identification of broad-leaved dock (*Rumex obtusifolius* L.) on grassland by means of digital image processing. *Precision Agriculture*, 7, 165–178.
- Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, 285–339.
- Govaerts, R. (2001). How many species of seed plants are there? *Taxon*, 50, 1085–1090.
- Govaerts, R., Wilkin, P., Raz, L., & Téllez-Valdés, O. (2010). World checklist of Dioscoreaceae, The Board of Trustees of the Royal Botanic Gardens, Kew.
- Gubatz, S., Dercksen, V. J., Brüß, C., Weschke, W., & Wobus, U. (2007). Analysis of barley (*Hordeum vulgare*) grain development using three-dimensional digital models. *The Plant Journal*, 52, 779–790.
- Gu, X., Du, J.-X., & Wang, X.-F. (2005). Leaf recognition based on the combination of wavelet transform and Gaussian interpolation. *Advances In Intelligent Computing*, 3644, 253–262.
- Haigh, A., Wilkin, P., & Rakotonasolo, F. (2005). A new species of *Dioscorea* L. (Dioscoreaceae) from Western Madagascar and its distribution and conservation status. *Kew Bulletin*, 60, 273–281.
- Hearn, D. J. (2009). Shape analysis for the automated identification of plants from images of leaves. *Taxon*, 58, 934–954.
- Hemming, J., & Rath, T. (2001). PA-Precision agriculture: Computer-Vision-based weed identification under field conditions using controlled lighting. *Journal of Agricultural Engineering Research*, 78, 233243.
- Hetherington, A. M., & Woodward, F. I. (2003). The role of stomata in sensing and driving environmental change. *Nature*, 424, 901–908.
- Hong, A., Chen, G., Li, J. L., Chi, Z. R., & Zhang, D. (2004). A flower image retrieval method based on ROI feature. *Journal of Zhejiang University-Science*, 5, 764–772.
- Hu, M. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8, 179–187.
- Huang, Q., Jain, A., Stockman, G., & Smucker, A. (1992). Automatic image analysis of plant root structures. In *Proceedings 11th IAPR international conference on pattern recognition, conference B: Pattern recognition methodology and systems* (Vol. II, pp. 569–572).
- Huang, Z.-K., De-Shuang, H., Du, J.-X., Quan, Z.-H., & Guo, S.-B. (2006). Bark classification based on Gabor filter features using RBPNN neural network. *International conference on neural information processing* (pp. 80–87). Springer.
- Huff, P. M., Wilf, P., & Azumah, E. J. (2003). Digital future for paleoclimate estimation from fossil leaves? Preliminary results. *Palaios*, 18, 266–274.
- Im, C., Nishida, H., & Kunii, T. L. (1999). Recognizing plant species by normalized leaf shapes. In *Proceedings of Vision interface* (pp. 19–21).
- Jensen, R. J., Ciofani, K. M., & Miramontes, L. C. (2002). Lines, outlines, and landmarks: Morphometric analyses of leaves of *Acer rubrum*, *Acer saccharinum* (Aceraceae) and their hybrid. *Taxon*, 51, 475–492.
- Kirchgessner, N., Scharr, H., Schurr, U. (2002). Robust vein extraction on plant leaf images. In *2nd IASTED international conference visualisation, imaging and image processing*.

- Kuhl, F. P., & Giardina, C. R. (1982). Elliptic Fourier features of a closed contour. *Computer graphics and image processing*, 18, 236–258.
- Lee, C., & Chen, S. (2006). Classification of leaf images. *International Journal of Imaging Systems and Technology*, 16, 15–23.
- Lexer, C., Joseph, J., van Loo, M., Prenner, G., Heinze, B., Chase, M. W., et al. (2009). The use of digital image-based morphometrics to study the phenotypic mosaic in taxa with porous genomes. *Taxon*, 58, 349–364.
- Li, Y., Chi, Z., & Feng, D. D. (2006). Leaf vein extraction using independent component analysis. *IEEE International Conference on Systems, Man and Cybernetics, IEEE* (pp. 3890–3984). IEEE.
- Ling, H., & Jacobs, D. W. (2007). Shape classification using the inner distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29, 286–299.
- Lipske, M. (2008). New electronic field guide uses leaf shapes to identify plant species. *Inside Smithsonian Research Winter*.
- Liu, J., Zhang, S., & Deng, S. (2009). A method of plant classification based on wavelet transforms and support vector machines. *Emerging Intelligent Computing Technology and Applications*, 5754, 253–260.
- MacLeod, N. (1999). Generalizing and extending the eigenshape method of shape space visualization and analysis. *Paleobiology*, 107–138.
- MacLeod, N., Benfield, M., & Culverhouse, P. (2010). Time to automate identification. *Nature*, 467, 154–155.
- Ma, W., Zha, H., Liu, J., Zhang, X., Xiang, B. (2008). Image-based plant modeling by knowing leaves from their apexes. In: 19th International conference on pattern recognition (pp. 1–4).
- McLellan, T., & Ender, J. A. (1998). The relative success of some methods for measuring and describing the shape of complex objects. *Systematic Biology*, 47, 264–281.
- Meade, C., & Parnell, J. (2003). Multivariate analysis of leaf shape patterns in Asian species of the Uvaria group (Annonaceae). *Botanical Journal Of The Linnean Society*, 143, 231–242.
- Melville, R. (1937). The accurate definition of leaf shapes by rectangular coordinates. *Annals of Botany*, 1, 673–679.
- Melville, R. (1939). Contribution to the study of British elms II. *Journal of Botany*, 77, 138.
- Mokhtarian, F., & Abbasi, S. (2004). Matching shapes with self-intersection: Application to leaf classification. *IEEE Transactions Image Processing*, 13, 653–661.
- Mullen, R., Monokosso, D., Barman, S., Remagnino, P., & Wilkin, P. (2008). Artificial ants to extract leaf outlines and primary venation patterns. *Lecture Notes In Computer Science*, 5217, 251–258.
- Nam, Y., Hwang, E., & Kim, D. (2005). CLOVER: A mobile content-based leaf image retrieval system. In *Digital Libraries: Implementing Strategies and Sharing Experiences. LNCS* (vol. 3815, pp. 139–148). Berlin/ Heidelberg: Springer.
- Nam, Y., Hwang, E., & Kim, D. (2008). A similarity-based leaf image retrieval scheme: Joining shape and venation features. *Computer Vision And Image Understanding*, 110, 245–259.
- Natho, G. (1959). Variationsbreite und bastardbildung bei mitteleuropäischen birkenstippen. *Repertorium novarum specierum regni vegetabilis*, 61, 211–273.
- Neto, J., Meyer, G., Jones, D., & Samal, A. (2006). Plant species identification using elliptic Fourier leaf shape analysis. *Computers And Electronics In Agriculture*, 50, 121–134.
- Nilsback, M. E., Zisserman, A. (2007). Delving into the whorl of flower segmentation. In *Proceedings of the British machine vision conference* (Vol. 1, pp. 570–579).
- Oakley, D., & Falcon-Lang, H. J. (2009). Morphometric analysis of cretaceous (Cenomanian) angiosperm woods from the Czech Republic. *Review of Palaeobotany and Palynology*, 153, 375–385.
- Pan, J., He, Y. (2008). Recognition of plants by leaves digital image and neural network. In *2008 International conference on computer science and software engineering, Wuhan, China* (pp. 906–910).
- Park, J., Hwang, E., & Nam, Y. (2008). Utilizing venation features for efficient leaf image retrieval. *Journal of Systems and Software*, 81, 71–82.
- Patterson, C. (1982). Morphological characters and homology. In K. Joysey & A. Friday (Eds.), *Problems of phylogenetic reconstruction* (vol. 576, pp. 21–74). Academic Press.
- Pauwels, E. J., de Zeeum, P. M., & Ranguelova, E. B. (2009). Computer-assisted tree taxonomy by automated image recognition. *Engineering Applications Of Artificial Intelligence*, 22, 26–31.
- Plotze, R. d., Falvo, M., Padua, J. G., Bernacci, L. C., Vieira, M. L. C., Oliveira, G. C. X., et al. (2005). A new morphometric method: A study with Passiflora (Passifloraceae). *Canadian Journal Of Botany*, 83, 287–301.
- Prusinkiewicz, P., & Lindenmayer, A. (1990). *The algorithmic beauty of plants*. New York, NY, USA: Springer-Verlag New York, Inc.
- Prusinkiewicz, P. (2004). Art and science for life: Designing and growing virtual plants with L-systems. In *XXVI International Horticultural Congress* (pp. 15–28).
- Qi, L., Yang, Q., Bao, G., Xun, Y., Zhang, L. (2009). A dynamic threshold segmentation algorithm for cucumber identification in greenhouse. In *2nd International congress on image and signal processing* (pp. 1–4).
- Ramos, E., & Fernandez, D. S. (2009). Classification of leaf epidermis microphotographs using texture features. *Ecological Informatics*, 4, 177–181.
- Ray, T. S. (1992). Landmark eigenshape analysis: Homologous contours: Leaf shape in syngonium (Araceae). *American Journal of Botany*, 79, 69–76.
- Royer, D. L. (2001). Stomatal density and stomatal index as indicators of paleoatmospheric CO₂ concentration. *Review of Palaeobotany and Palynology*, 114, 1–28.
- Royer, D. L., & Wilf, P. (2006). Why do toothed leaves correlate with cold climates? Gas exchange at leaf margins provides new insights into a classic paleotemperature proxy. *International Journal of Plant Sciences*, 167, 11–18.
- Royer, D. L., Wilf, P., Janesko, D. A., Kowalski, E. A., & Dilcher, D. L. (2005). Correlations of climate and plant ecology to leaf size and shape: Potential proxies for the fossil record. *American Journal of Botany*, 92, 1141–1151.
- Rumpunen, K., & Bartsch, I. V. (2002). Comparison of differentiation estimates based on morphometric and molecular data, exemplified by various leaf shape descriptors and RAPDs in the genus *Chaenomeles*. *Taxon*, 51, 69–82.
- Schneider, C. K. (1912). *Handbuch der Laubholz-kunde*, Jena, Germany (vol. 1, pp. 367–389).
- Scotland, R. W., & Wortley, A. H. (2003). How many species of seed plants are there? *Taxon*, 52, 101–104.
- Soille, P. (2000). Morphological image analysis applied to crop field mapping. *Image and Vision Computing*, 18, 1025–1032.
- Song, Y., Wilson, R., Edmondson, R., Parsons, N. (2007). Surface modelling of plants from stereo images. In *Sixth international conference on 3-D digital imaging and modeling 3DIM '07* (pp. 312–319).
- Stace, C. A. (1992). *Plant Taxonomy and Biosystematics* (second ed.). Cambridge University Press.
- Stuessy, T. F. (2006). Principles and practice of plant taxonomy. In E. Leadley & S. Jury (Eds.), *Taxonomy and Plant Conservation* (pp. 31–44). Cambridge University Press.
- Teague, M. (1980). Image analysis via the general theory of moments. *J. Opt. Soc. Am.*, 70, 920–930.
- Teng, C. H., Kuo, Y. T., Chen, Y. S., & segmentation, Leaf (2009). its 3D position estimation and leaf classification from a few images with very close viewpoints. *Image Analysis and Recognition*, 937–946.
- Thiele, K. (1993). The holy grail of the perfect character: The cladistic treatment of morphometric data. *Cladistics*, 9, 275–304.
- Šeatović, D. (2008). A segmentation approach in novel real time 3D plant recognition system. In *Proceedings of the 6th international conference on computer vision systems* (pp. 363–372).
- Wang, Z., Chi, Z., & Dagan, F. (2003). Shape based leaf image retrieval. *Vision, Image And Signal Processing*, 150, 34–43.
- Wang, Z., Chi, Z., Dagan, F., & Wang, Q. (2000). Leaf image retrieval with shape features. *Advances In Visual Information Systems*, 1929, 41–52.
- Wang, X., Huang, D., Du, J., Xu, H., & Heutte, L. (2008). Classification of plant leaf images with complicated background. *Applied Mathematics and Computation*, 205, 916–926.
- Wheeler, Q. D. (2004). Taxonomic triage and the poverty of phylogeny. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359, 571–583.
- White, S., Feiner, S., Kopylec, J. (2006). Virtual vouchers: Prototyping a mobile augmented reality user interface for botanical species identification. In *IEEE Symposium on 3D user interfaces* (pp. 119–126).
- White, R., Prentice, H. C., & Vervijst, T. (1988). Automated image acquisition and morphometric description. *Canadian Journal of Botany*, 66, 450–459.
- Wilkin, P. (1999). A morphometric study of *Dioscorea quartiniiana* (Dioscoreaceae). *Kew Bulletin*, 54, 1–18.
- Wilson, E. O. (2003). The encyclopedia of life. *Trends in Ecology & Evolution*, 18, 77–80.
- Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y.-X., Chang, Y.-F., & Xiang, Q.-L. (2007). A leaf recognition algorithm for plant classification using probabilistic neural network. *IEEE international symposium on signal processing and information technology*. IEEE.
- Ye, L., & Keogh, E. (2009). Time series shapelets: A new primitive for data mining. *IEEE International Conference on Knowledge Discovery and Data Mining* (pp. 947–956). ACM.
- Yoshioka, Y., Iwata, H., Ohsawa, R., & Ninomiya, S. (2004). Analysis of petal shape variation of *Pimula sieboldii* by elliptic Fourier descriptors and principal component analysis. *Annals of Botany*, 94, 1657–1664.
- Young, J. P., Dickinson, T. A., & Dengler, N. G. (1995). A morphometric analysis of heterophyllous leaf development in *Ranunculus flabellaris*. *International Journal Of Plant Species*, 156, 590–602.
- Zarinkamar, F. (2007). Stomatal observations in dicotyledons. *Pakistan Journal of Biological Sciences*, 10, 199–219.
- Zelditch, M., Swiderski, D., & Fink, W. (2000). Discovery of phylogenetic characters in morphological data. In J. Wiens (Ed.), *Phylogenetic Analysis of Morphological Data* (pp. 37–83). Washington and London: Smithsonian Institution Press.
- Zeng, G., Birchfield, S. T., & Wells, C. E. (2008). Rapid automated detection of roots in minirhizotron images. *Machine Vision and Applications*, 21, 309–317.