Using the Hausdorff Dimension Analysis of Tree Branching Patterns as a Method of Species Differentiation and Identification

Professor John Boccio
Andreas Bastian

Introduction

Fractal structures are found at many scales in Nature and are especially frequently seen in biological systems. Everything from circulatory systems to the structures of sea sponges to the flowering bodies of common vegetables manifest fractal geometry. As Felix Hausdorff discovered, fractals can be described as having non-integer dimension. In this paper, I will explore the concept of fractal dimension and its relationships to the branching structures of trees in hopes of applying fractal dimension as a tool for deciduous tree species differentiation and identification.

Theory

Fractals

Fractals encompass a relatively simple set of properties that appear in a hugely diverse variety of systems. A structure is often called "fractal" if it has the following properties:

1. Scale invariance (infinite detail)
2. Self-similarity
3. Non-integer Hausdorff dimension

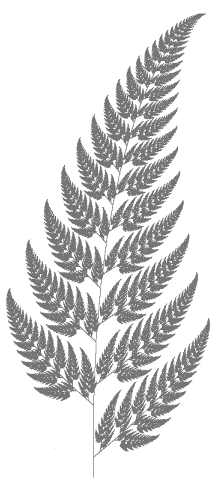
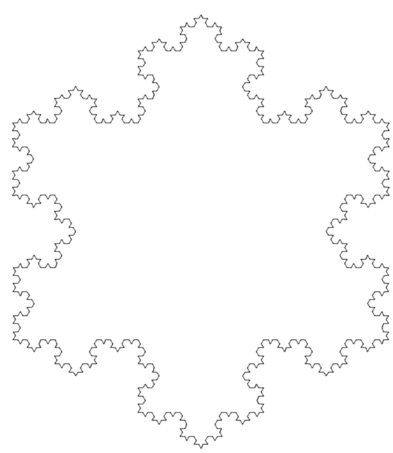

Figure 1. Two common examples of fractals: the fractal fern [1] and the Koch snowflake [2]. Each has scale invariance, self-similar patterns, and a non-integer Hausdorff dimension.

Hausdorff Dimension

The Hausdorff dimension is a piece of information that can be extracted from a careful geometric analysis of a shape. Beginning with a one-dimensional example, if one were to try to cover a line of length *l* in one-dimensional space with a "measuring line" of the same length l, it would only take one line of this length. If instead we used measuring lines of length $l/_2$, it would take two of these lines to cover the line of interest. If we continue to use smaller and smaller measuring lines of length $l/_{2^m}$, to cover the line of interest, we need 2^{m} lines to do so. The length of one of the 2^{m} measuring lines needed to cover a line of length *l* can be defined as

$$\delta_m = \frac{l}{2^m}$$

Using this definition, the number of measuring lines needed to measure a line of length *l* is

$$N(\delta_m) = \frac{l}{\delta_m} = \left(\frac{l}{\delta_m}\right)^1$$

In two dimensions, rather than using a measuring line, we can use a measuring square box to measure the two-dimensional analog of length: area. Beginning with a box with side length *l*, it takes only one measuring box of side length l and area l^2 to describe the area of the box of interest. If the measuring boxes are reduced to having side length $l/_2$, each has an area of $l^2/_4$ or $(l/_2)^2$, requiring a total of four boxes to describe the area of the box. Expanding upon this, 2^m boxes of area $(l/_{2^m})^2$ are needed to describe the area of the box of interest. The number of boxes required can be described as a function of the area of the measuring boxes:

$$N(\delta_m) = \left(\frac{l}{\delta_m}\right)^2$$

Continuing the trend observed for the one- and two-dimensional cases, the expression for the number of "measuring boxes" necessary can be generalized to any dimension *D*:

$$N(\delta_m) = \left(\frac{l}{\delta_m}\right)^D$$

Solving for the dimension *D* of the shape being analyzed via this method yields the equation

$$D = \lim_{m \to \infty} \frac{\log(N(\delta_m))}{\log(l) + \log(\frac{1}{\delta_m})}$$

As the number of measuring boxes *m* approaches infinity, *l* approaches zero, as does $\log(l)$:

$$D = \lim_{m \to \infty} \frac{\log(N(\delta_m))}{\log\left(\frac{1}{\delta_m}\right)}$$

The dimension of an object can be described as the ratio of the log of number of *n*-dimensional measuring boxes needed to cover it divided by the log of the *n*-dimensional analog to box area.

As is illustrated in figure 2, the number N of n-dimensional boxes necessary to describe a Euclidian object scales as the power of the ratio of the measuring box "length" and the length of the measured object.

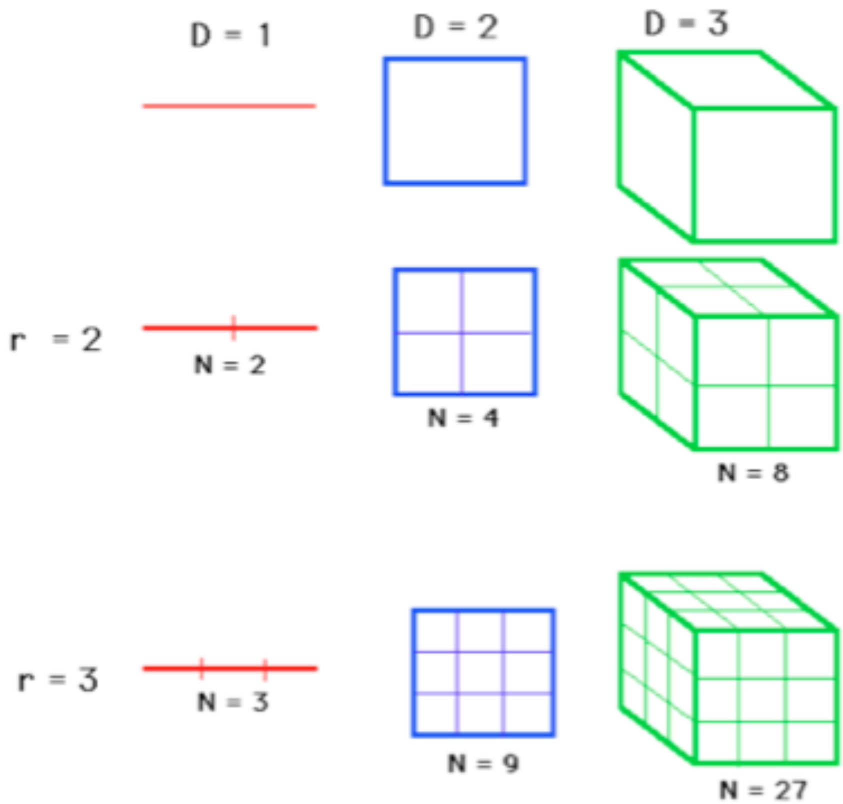

Figure 2. The relationship between the box length ratio and the number of boxes necessary to describe an object. (Image: John Boccio).

This can be written as $N = r^D$, where r is the ratio of measuring box "length" to the overall "length" of the object of interest: $r = {l_{box}}/{l_{object}}$. Solving for D, we get the following equation:

$$D = \frac{\log(N)}{\log(r)}$$

This definition of dimension yields integer values for Euclidian shapes and non-integer values for fractal sets. In a plot of log(N) vs. log(r), D is the slope of the line resulting from iterative reduction in the "length" of the measuring boxes.

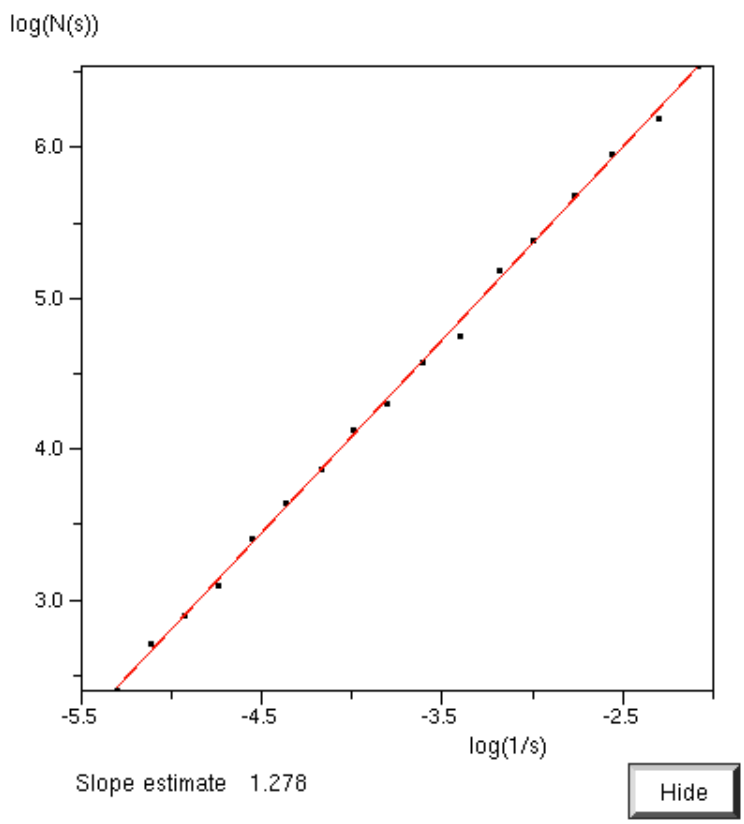

Figure 3. Plot of log(N) vs. log(r), illustrating the dimension D as the ratio of these two quantities.

Methods

Image preparation

I took several measures to eliminate inconsistencies between image samples to ensure clean data output from the FDC UNIX program. First, I took all of my source images with the same camera within a twenty-minute timeframe under constant daylight conditions. I selected four species for my study: *Quercus shumardii, Quercus bicolor, Tilia tomentosa,* and *Carya ovata.* I scaled all the source images to 640x480 pixels. I ran these large images as well as 200 pixel square subsets of these images through the FDC program to extract fractal dimensions. Additionally, I converted the large source images to black and white via thresholding and ran these through the FDC program as well. Some manual adjustment was necessary to achieve clear structure definition.

Color versus black and white images

The same image yielded different Hausdorff dimension numbered when modified to be a high-contrast black and white image versus when it was left in its natural color state. The FDC program had no trouble correctly identifying the fractal structure of the tree limbs in both pictures, as can be seen below:

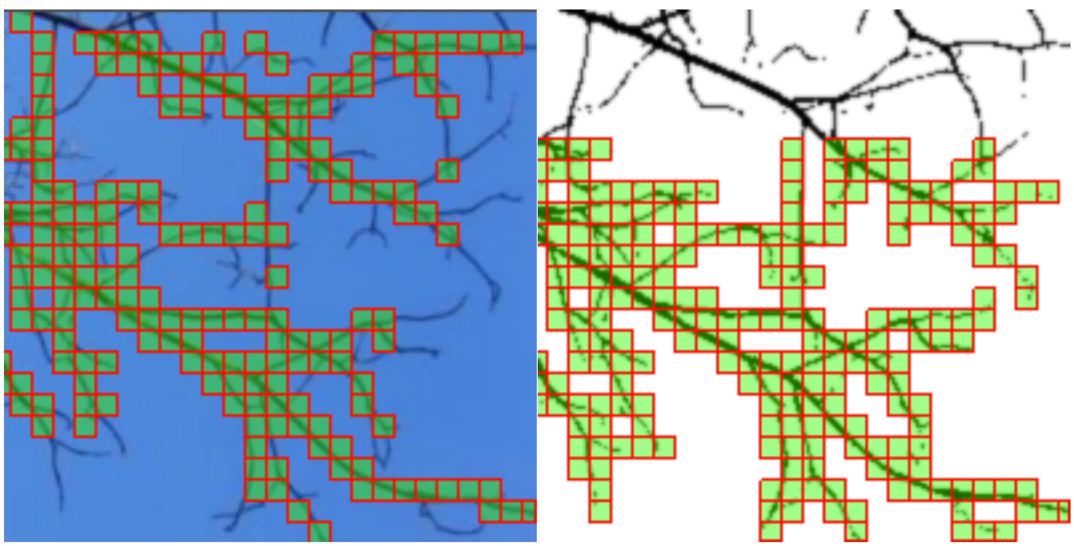

Figure 4. Comparison of FDC performance using a full-color source image and a black and white source image.

The FDC program is able to identify the large-scale features of the tree branch structure in the full-color image, but it has trouble with the finer features. However, in the black and white image, the FDC program is able to capture and analyze the entire structure of the branching pattern. Based on this observation, I decided to render all source images for the FDC program in black and white to ensure that the

Hausdorff dimension numbers that I calculated accurately reflected the branching structure.

Variation of the Hausdorff dimension for an individual image

The FDC program does not necessarily return the same result after each iteration of box counting. As can be seen in the chart below, the fractal dimension returned for each run of the FDC program can vary by a little over 1%.

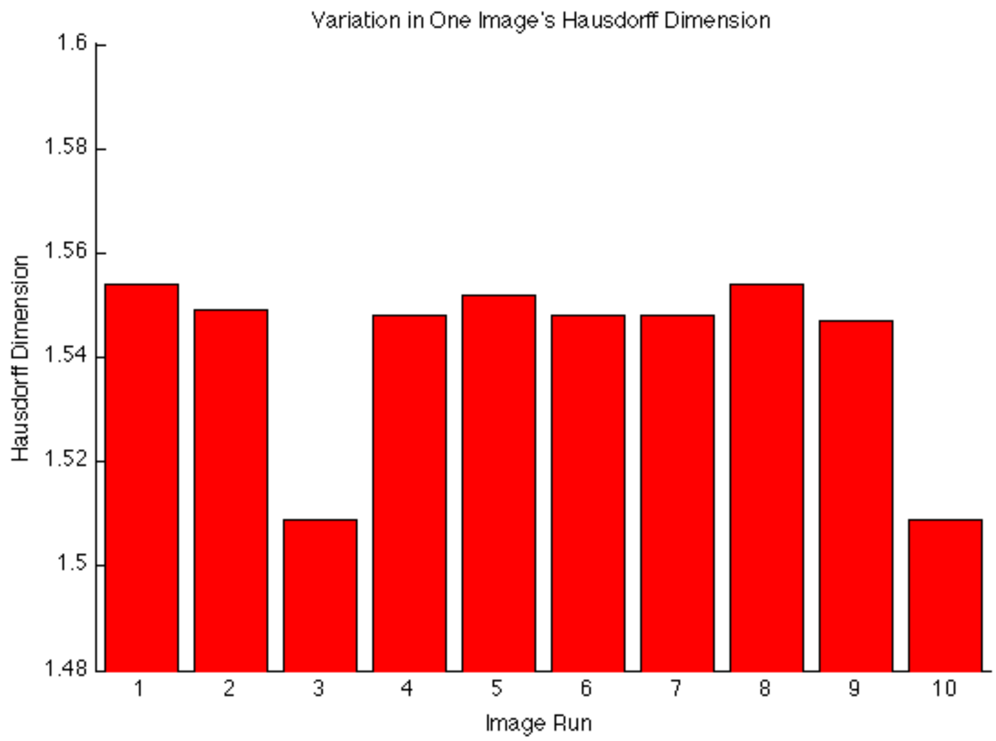

Figure 5. Ten runs of the FDC analysis program on the same image.

Variation of source image size

Image size is a critical factor in determining the Hausdorff dimension of a given branching pattern. I ran the FDC program on 200 pixel square images and 640x480 pixel images to determine the relationship between image size and fractal dimension.

Results

The differences between the mean Hausdorff dimensions of the four tree species I examined are significant to varying degrees, as illustrated in the following tables and charts.

	Q. shumardii	Q. bicolor	T. tomentosa	C. ovata
Image 1:	1.71	1.694	1.776	1.747
Image 2:	1.745	1.679	1.825	1.746
Image 3:	1.759	1.691	1.809	1.776
Image 4:	1.736	1.697	1.828	1.775
Image 5:	1.811	1.698	1.749	1.771
Mean:	1.7522	1.6918	1.7974	1.763
St. Dev.	0.037412565	0.007661593	0.034033807	0.015182226
St. Dev. (%)	2.13517662	0.452866327	1.893502099	0.861158619

Table 1. Mean Hausdorff dimensions of four different tree species' branching patterns.

	Q. shumardii	Q. bicolor	T. tomentosa	C. ovata
Q. shumardii	-	P=0.0077	P=0.0807	P=0.5663
Q. bicolor	-	-	P=0.0001	P=0.001
T. tomentosa	-	-	-	P=0.0729
C. ovata	-	-	-	-

Key:	Not significant	Somewhat significant	Very significant

Table 2. Two-tailed P-values from unpaired t-tests between Hausdorff dimensions extracted from large black and white images.

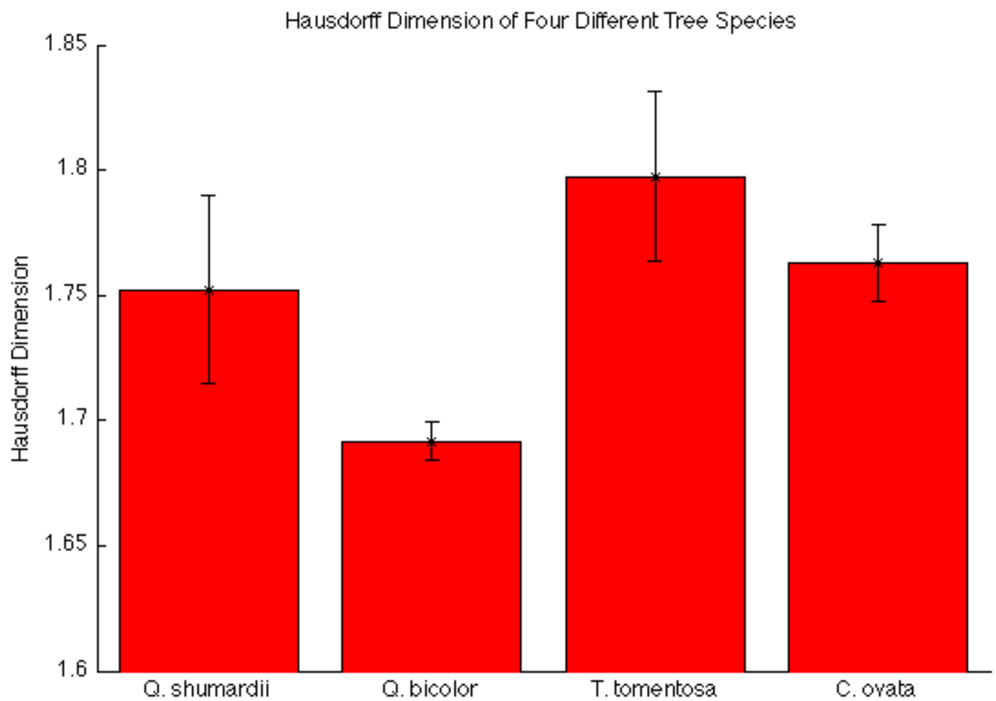

Figure 6. The mean Hausdorff dimensions and associated standard deviations for large, black and white images of branching patterns.

	Q. shumardii	Q. bicolor	T. tomentosa	C. ovata
Q. shumardii	-	P=0.0154	P=0.2821	P=0.5799
Q. bicolor	-	-	P=0.0012	P=0.0012
T. tomentosa	-	-	-	P=0.0517
C. ovata	-	-	-	-

Key:	Not significant	Somewhat significant	Very significant

Table 3. Two-tailed P-values from unpaired t-tests between Hausdorff dimensions extracted from large color images.

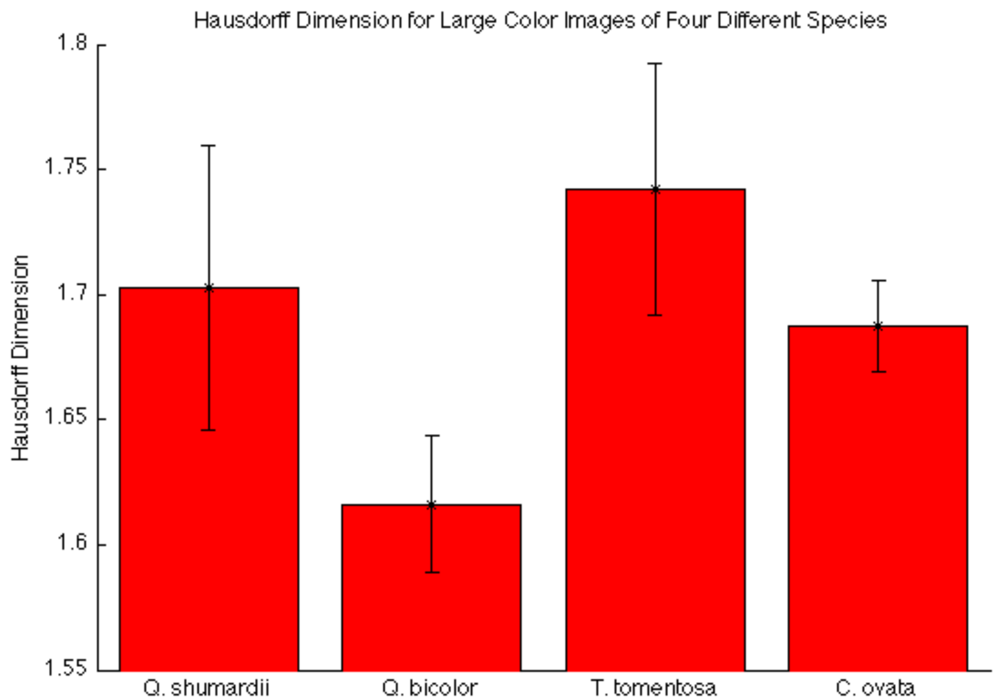

Figure 7. Mean Hausdorff dimensions for full-color, full-sized images.

	Q. shumardii	Q. bicolor	T. tomentosa	C. ovata
Q. shumardii	-	P=0.2969	P=0.5478	P=0.5615
Q. bicolor	-	-	P=0.7336	P=0.5519
T. tomentosa	-	-	-	P=0.8763
C. ovata	-	-	-	-

Key:	Not significant	Somewhat significant	Very significant

Table 4. Two-tailed P-values from unpaired t-tests between Hausdorff dimensions extracted from 200 pixel square color images.

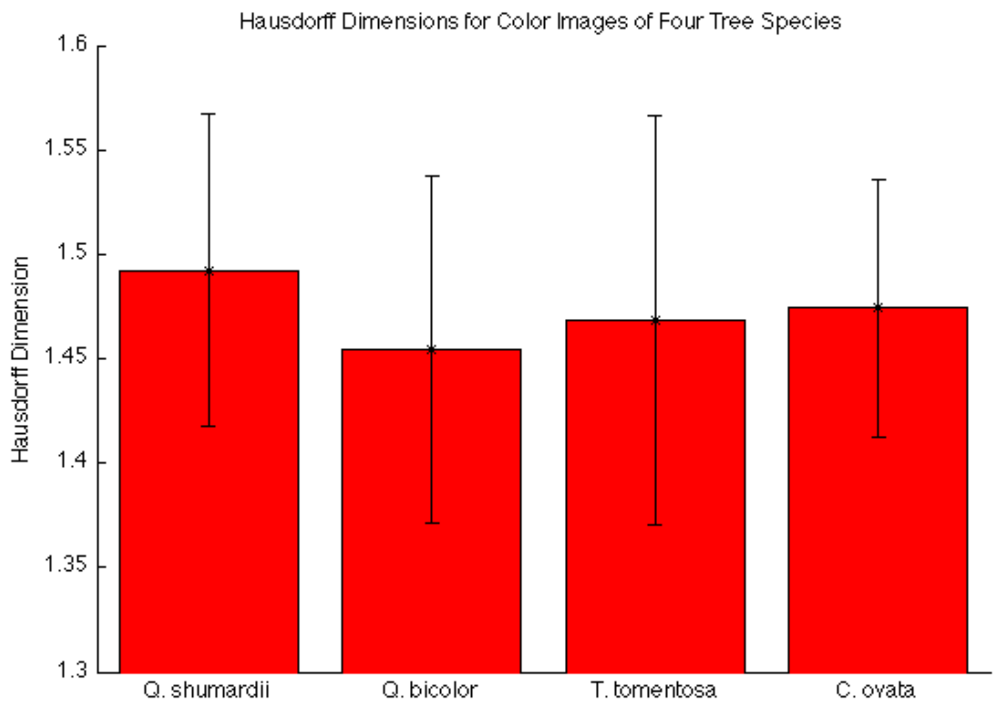

Figure 8. The mean Hausdorff dimensions collected from small, full-color images of tree branching patterns.

Discussion

The results indicate suggest several relationships between branching structure, image characteristics, and fractal dimension. First, different species of tree appear to have unique Hausdorff dimensions. A species' Hausdorff dimension can be used to differentiate it from some other species, but this property cannot be used for reliable species identification because different species can yield mean Hausdorff dimensions that are impossible to differentiate. However, in some cases, the Hausdorff can be used to differentiate two species, provided that their Hausdorff dimensions are distinguishable to begin with, or to narrow the pool of possible identifications.

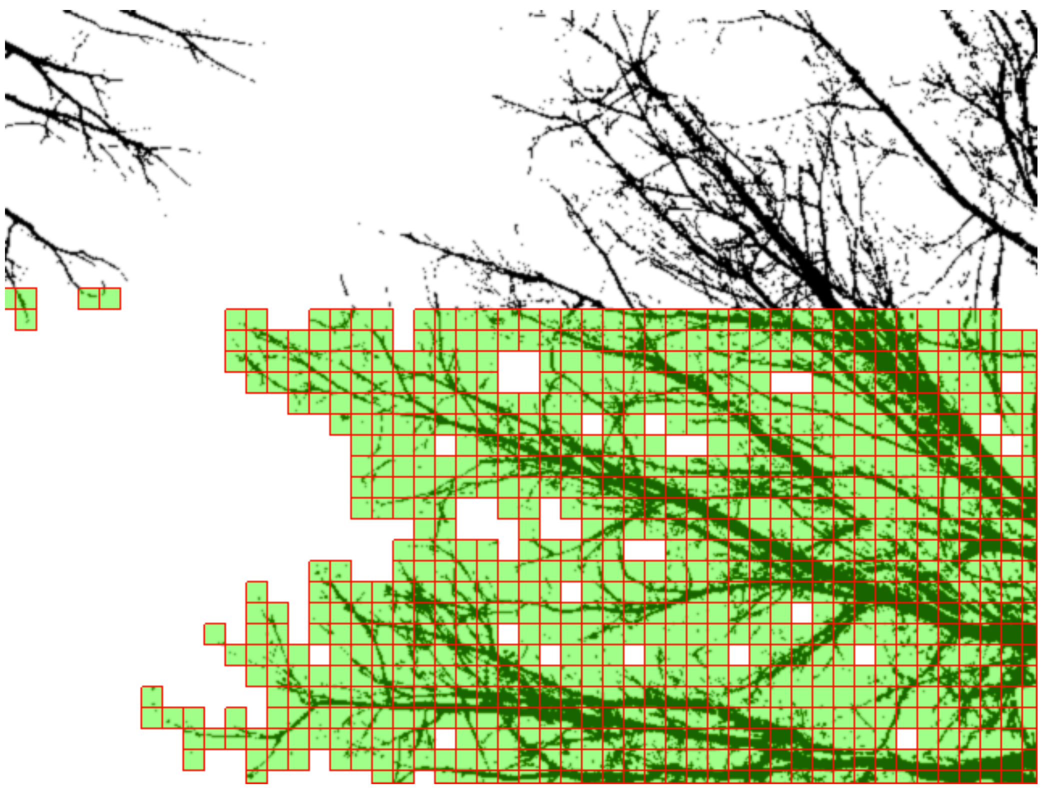

Figure 9. Large, high-contrast black and white images yielded far more consistent and statistically distinguishable results.

The properties of the input image are remarkably important. As can be seen in the comparison between small, color images and large black and white images, image properties can make the difference between statistically distinguishable means and statistically insignificant differences. Black and white images yield far more reliable results, as do larger images. Images with both attributes are the best candidates for Hausdorff dimension analysis with the FDC program.

Conclusions

The Hausdorff dimension is a powerful tool for fractal analysis and can, in some cases, be used to differentiate tree species based on their branching patterns. This technique could perhaps be scaled and applied to other fractal forms found in Nature as a tool to facilitate further understanding of our natural world.

Image Credits

[1] “koch,” n.d., from http://www.spaennare.se/FRACTAL/koch.jpg
[2] “fern,” n.d., from http://www.spaennare.se/FRACTAL/fern.jpg