Borrower: RAPID:NAM

Lending String:

Patron:

Journal Title: IFAR journal / International Foundation for Art Research.

Volume: 8 **Issue:** 2
Month/Year: 2005**Pages:** 12-

Article Author: Rockmore, Daniel,

Article Title: A Digital Technique for Authentication in the Visual Arts.

Imprint:

ILL Number: -6802087

ODYSSEY ENABLED

Charge
Maxcost:

Shipping Address:
NEW: Main Library
same

Fax:
Ariel: 169.226.84.218
Email:

A DIGITAL TECHNIQUE FOR AUTHENTICATION IN THE VISUAL ARTS

DANIEL ROCKMORE, SIWEI LYU, AND HANY FARID*

Can mathematics help the connoisseur? Dartmouth College researchers present a mathematical measure of an artist's style that may yield a new quantitative tool for the authenticator's kit.

EDITOR'S NOTE:

The process of attributing or authenticating a work of art, which IFAR has been involved with for 36 years, has never been an exact science, but rather, depending on the work, a combination of connoisseurship, scholarly documentation (including provenance), and examination of the physical properties of the work. In recent years, however, several mathematical and scientific methodologies that take advantage of sophisticated new computer programs to analyze elements of an artist's style have been introduced. Although none of them have been tested on a large enough universe of artworks to convince skeptics, or even proponents, several do hold out the promise of adding valuable new tools to the authenticator's arsenal. The following article is the first of several that *IFAR Journal* will publish by scientists and mathematicians working in the new field, not as an endorsement *per se* but to introduce our readers to these new methodologies.

INTRODUCTION

It probably wasn't long after people began paying money for art that a lucrative business in forgery was born. And it probably wasn't too much later when techniques for detecting art forgeries emerged. The early authentication techniques remain preeminent. By and large these are based on connoisseurship – the discerning "eye" of an expert steeped in the work and life of the artist in question, whose opinion may be informed by the *catalogue raisonné*, the authoritative documentation of the artist's œuvre. Other desiderata may include provenance, which, ideally, could trace the work back to the artist's studio or circle or his known early collectors. Detailed analysis of a signature, if present, may also be instructive.[1]

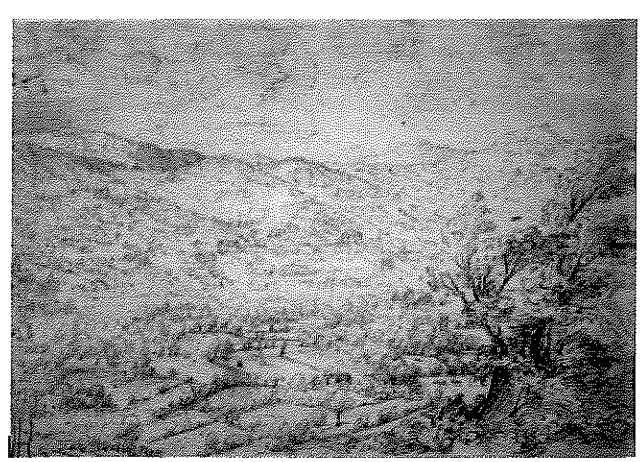

FIGURE 1. *Mountain Landscape with Ridge and Valley* (c.1552), Herzog Anton Ulrich Museum Braunschweig, Kupferstichkabinett Z. 381, by Pieter Bruegel the Elder.

"Is the hand of the artist implicit in the numerical or digital representation of a work of art? In this *IFAR Journal* article we suggest that the answer. . . is 'yes'."

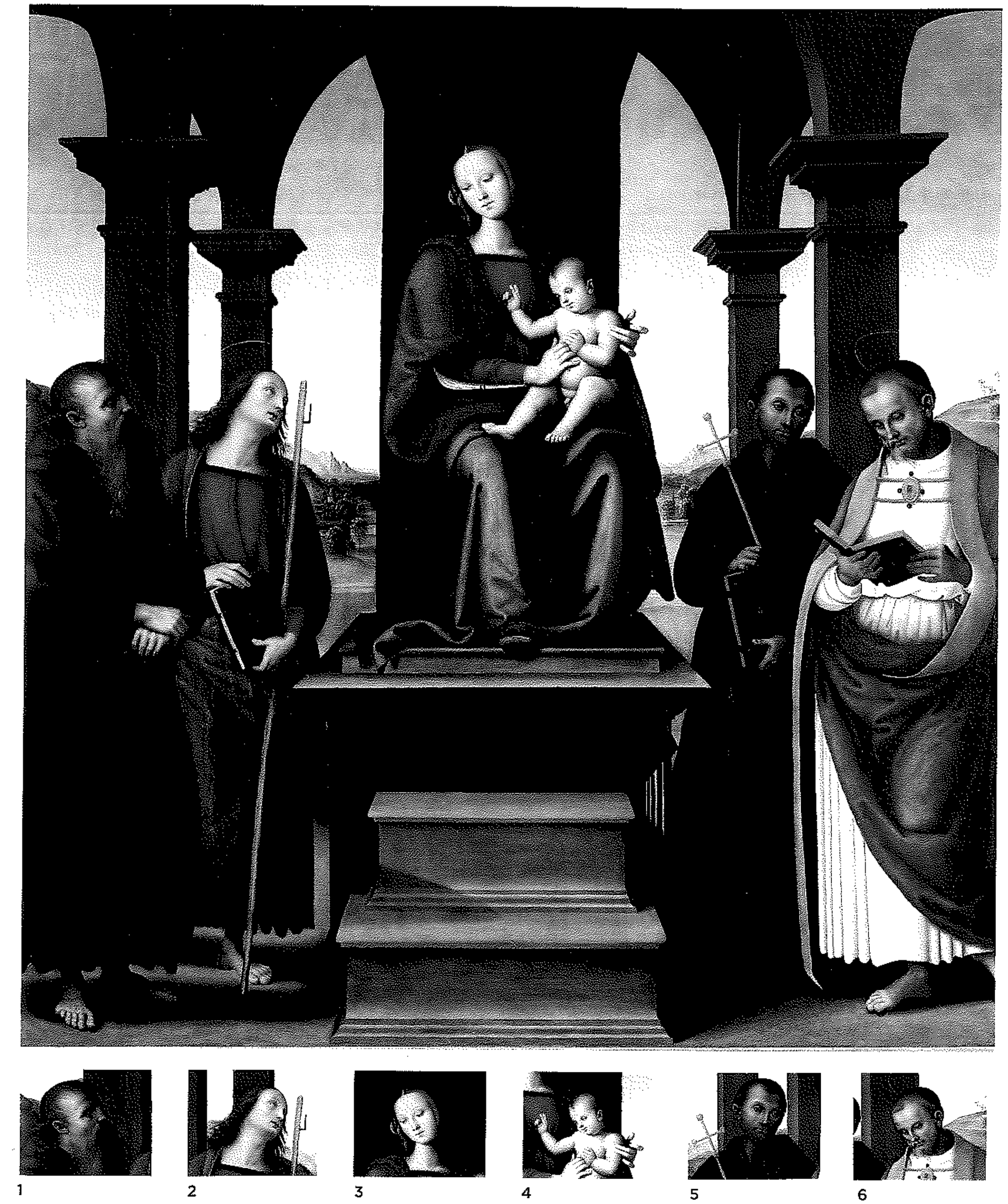

FIGURE 2. *Madonna and Child with Saints*, c.1500. Oil and tempera on panel, 175.2 x 172 cm, attributed to Perugino and assistants. Hood Museum, Dartmouth College, Hanover, New Hampshire. Details #1-6 show the faces in the paintings from left to right, starting with the bearded figure of Saint Anthony Abbot on the far left.

"Our work bears some similarity to computational approaches used to help determine literary authorship . . ."

However, in this day and age when practically any work of art can be and often is digitized and thereby transformed to a set of numbers resident within a computer, it is natural to wonder if mathematical tools might be brought to bear on the problem of authentication. That is, is the hand of the artist implicit in the numerical or digital representation of a work of art? In this *IFAR Journal* article, we suggest that the answer to such a question is "yes." We describe a new computational method for analyzing prints, drawings and paintings, which, preliminary research suggests, is useful for authentication. This approach builds a statistical model of an artist from the scans of a set of authenticated works, against which new works are then compared. In our initial investigations we have applied our tools to look for distinctions both *across* different works – e.g., the drawings of Pieter Bruegel the Elder (Fig. 1) – as well as *within* an individual work – e.g., the faces of the six figures in Perugino's *Madonna and Child* (Fig. 2). With Bruegel, we were successful in distinguishing eight authenticated drawings from works by various imitators (based on the attributions of the current catalogue raisonné).[2] While with the Perugino, our analysis found evidence of four different painters, a result that is in line with at least some expert opinion.[3]

While our techniques draw from relatively modern mathematical research, they do not by any means mark the first time mathematical methods have been brought to bear on the problem of authentication. Preeminent are the tried and true tools for dating a work via materials analysis through a basic application of the mathematics of *differential equations*. Of greater mathematical interest is the relatively recent application of ideas from the discipline of *machine learning* (an outgrowth of work on artificial intelligence) to the analysis and classification of craquelure and signatures. In the special case of the drip paintings of Jackson Pollock, the mathematics of fractals appears to produce a reliable quantitative stylistic signature, while also contributing to a discussion of the evolution of Pollock's aesthetic.[4]

Our work bears some similarity to computational approaches used to help determine literary authorship, a field which has come be known as *stylometry*[5]. Indeed, as more and more art collections are digitized (see, e.g., the ArtSTOR Project[6]), mathematical and computational techniques should begin to play a more important role in the analysis of visual artwork. We view the methodology outlined below as a potential new non-invasive and non-destructive computational technique for the authenticator's toolkit.

SOME RESULTS

Let us begin by showing some results. For this it suffices to know that from a corpus of high resolution digitized images of the original works, our procedure automatically produces a list of numbers for each image that ultimately is whittled down to a summary given by three numbers per picture (in the case of the Bruegels and their imitations) or three numbers per detail (in the case of the Perugino). As part of the preparation, we convert these color images to a grayscale version.[7] This enables us to attend to attributes of line. Each grayscale digital image comprises just over 4 million pixels, which are in turn represented in the computer as just over 4 million numbers (one number per pixel). Each number represents the intensity of light at the associated pixel position. This numerical encoding of the image is used as input for a sequence of mathematical steps (briefly outlined later in this article) that ultimately produces the list of three numbers that serves as the numerical summary of the original image. As any list of three numbers naturally corresponds to a point in the Cartesian representation of three-dimensional space,[8] each image (picture or detail, depending on the situation) corresponds to a geometric point in space. The degree of similarity between images then corresponds to the distance between associated points.

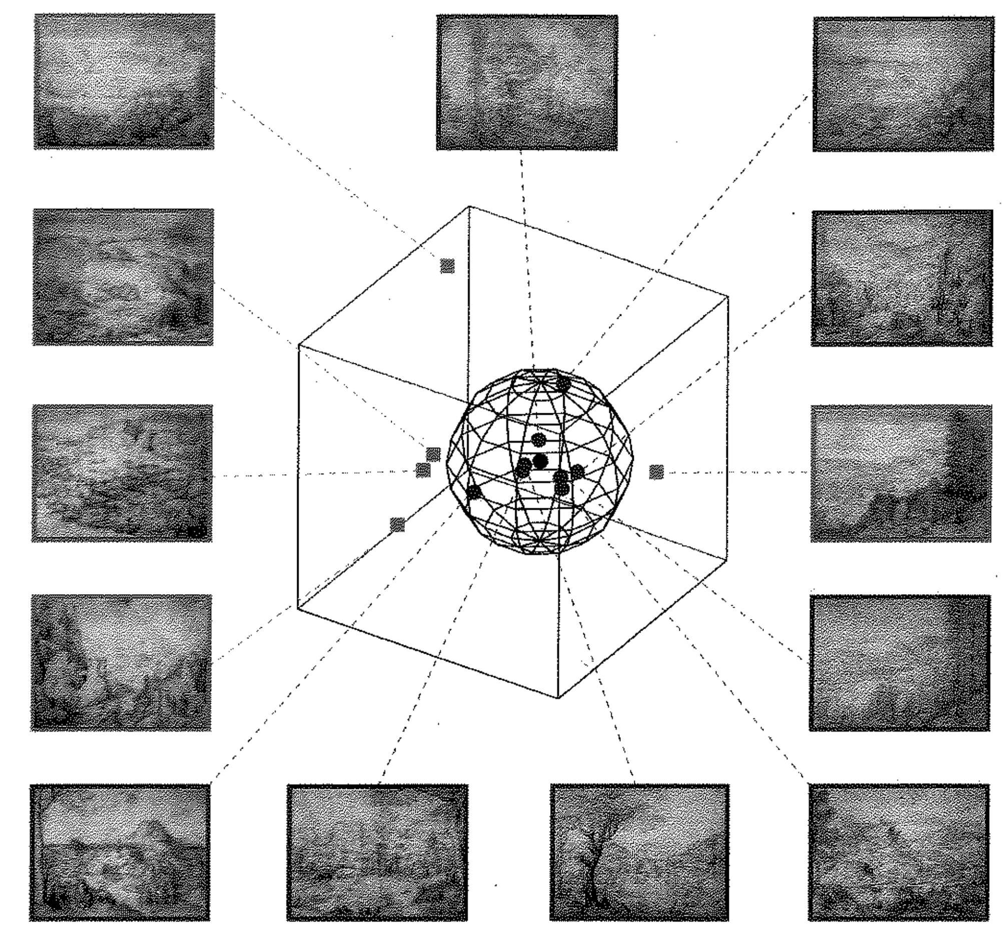

FIGURE 3. Results of analyzing two sets of authentic Bruegel drawings and imitations (13 works). The blue circles represent authentic Bruegel drawings, and the red squares represent imitations. The imitations lie significantly (in a statistical sense) outside of the bounding sphere of authentic drawings.

INVESTIGATION #1: Bruegel—or Not?

The Flemish painter and draftsman, Pieter Bruegel the Elder (1525/30-1569), was among the greatest artists of the sixteenth century. Over time he acquired many imitators, some undoubtedly simply eager to work in the style of the great master and others surely hoping to pass off their work as Bruegel's for monetary gain. Some of these followers and forgers were sufficiently expert that, after being unmasked (or discovered), they became famous in their own right (e.g., Jacob Savery).

"In one experiment we considered a set of thirteen landscapes by Bruegel and five acknowledged Bruegel imitations."

The authentication of Bruegel's work continues to be a subject of great interest. Many drawings formerly attributed to Bruegel are now attributed to others, and what was at the turn of the twentieth century a body of work of about one hundred drawings has over time been trimmed to approximately fifty.[9] Motivated by the Metropolitan Museum of Art's 2001 exhibit "Pieter Bruegel the Elder: Drawings and Prints" (curated by Nadine M. Orenstein), which contained Bruegel works possessing secure attribu-

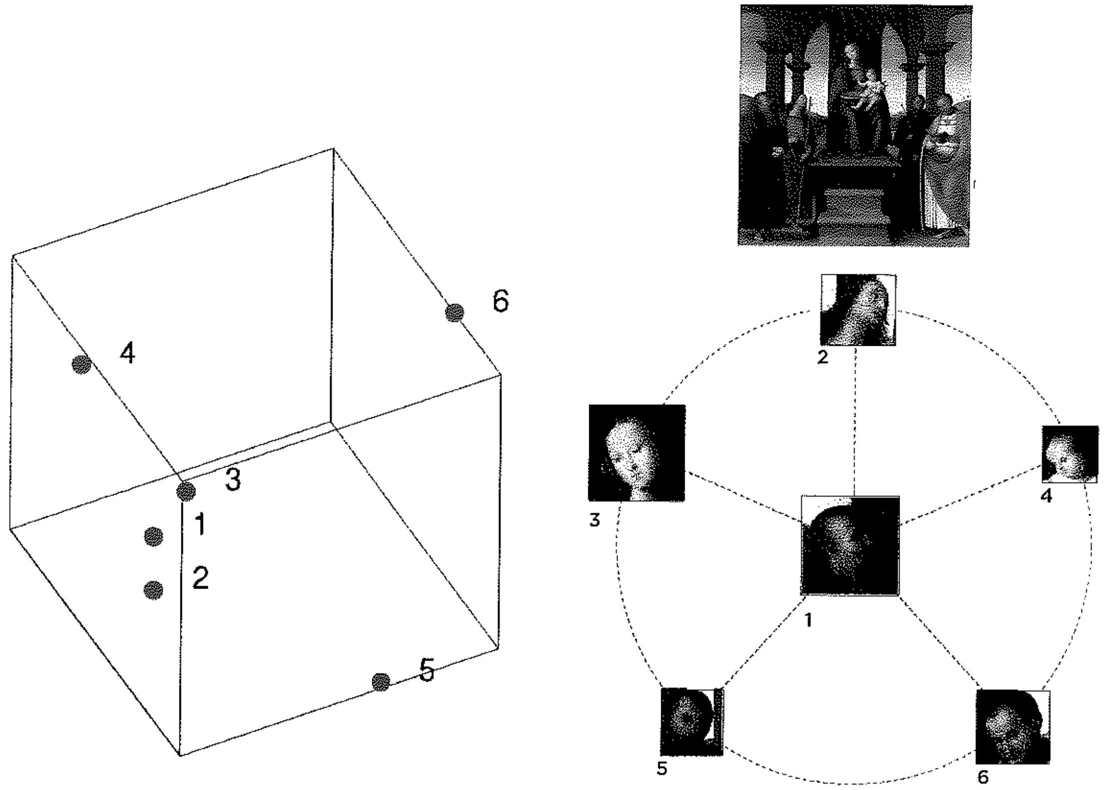

FIGURE 4. Results of analyzing Perugino's *Madonna and Child.* The numbered data points in the cube correspond to the six faces indicated on the right (the numbering is from left to right of the painting starting with the bearded figure; see also Fig. 2). Note how the three leftmost faces (1-3) cluster, while the remaining faces are distinct. This clustering pattern suggests the presence of at least four distinct hands.

tion as well as others that are acknowledged imitations, we investigated the question of whether or not a mathematical comparison of the images would support some of the experts' attributions. In one experiment we considered a set of thirteen landscapes consisting of eight secure drawings by Bruegel and five acknowledged Bruegel imitations.[10] Our working data was obtained by digitally scanning 35mm color slides at a resolution of 2400 dots per inch (dpi).[11] After scanning, these color images were cropped to a central square region, 2048 pixels on a side. Finally, the mathematical (wavelet) analysis was applied to the cropped images.

The result of doing this for Bruegel and Bruegel-like images in our two sets of drawings is shown in Figure 3. Therein, the blue dots mark the numerical summaries of the authenticated Bruegels, and the red dots mark the imitations. Thumbnail images of the drawings surround the borders. Notice that in each case, the authentic Bruegels are closely clustered together and are well separated from the imitations. In fact, this clustering is "statistically significant" in the sense that the distance of this group of points from the others is of a greater extent than would be attributed to mere chance. The wire-frame sphere surrounding the Bruegel points has been drawn only to emphasize that the analysis produces a statistic by which the authentic Bruegels do in fact cluster together while simultaneously exhibiting a collective difference from each imitation. This suggests that the wavelet analysis can be useful for authentication. In the authentication scenario, presumably we would be working (as in this situation) with a body of known secure works against which the questioned work would be compared.

INVESTIGATION #2: Perugino — The Problem Of Determining *How Many Hands*?

Pietro di Cristoforo Vannucci (Perugino) (1450–1523) is well known as a portraitist and a fresco painter, but is, perhaps, best known for his altarpieces. By the 1490s, Perugino, a prolific artist, maintained workshops in both Florence and Perugia. As was often the case for a large work by a great Renaissance master, it is likely that Perugino only painted a portion of the *Madonna and Child* (Fig. 2; Hood Museum, Hanover, NH) while apprentices did the rest. To this end, we wondered if our technique could uncover mathematical differences among the faces of the individual characters in the work in order to help answer the question: *how many hands contributed to the work?*

Our technique focuses on the linear elements of the work. Thus, in order to best pick up the fine brushwork, we decided to photograph the painting using a large-format camera (providing an 8×10 inch transparency) and then drum-scan the image to obtain a color 16,852 × 18,204 pixel image. This results in a digital image that is approximately 300 megapixels, and is thus at a resolution that is approximately 60 times greater than that obtained by a standard digital camera. We chose to compare the facial regions of the six characters, which were cropped from the full image. Following the technique used in the Bruegel analysis, the pixel values for each of the facial details were then fed into the computational process. The result is a list of three numbers for each facial region, displayed graphically in Figure 4. The six faces are numbered from left to right as they appear in the painting. Note how points 1, 2, and 3 cluster, while the remaining points are distinct. This clustering pattern suggests the presence of at least four distinct hands with one hand responsible for the faces on the three leftmost figures. This is consistent with the views of some art historians, such as Bart Thurber (Curator at Dartmouth's Hood Museum), although not all. For example, Laurence Kanter, Curator-in-Charge of the Robert Lehman Collection at New York's Metropolitan Museum of Art, believes that the absence of color data (due to our decision to perform the analysis in grayscale rather than color) represents a significant omission.[12]

"... It is likely that Perugino only painted a portion of the *Madonna and Child* ... We wondered if our technique could uncover mathematical differences among the faces ..."

THE PROCESS: An Algorithm For Art Authentication

So, how was this accomplished? How is it that the digital representations of each of these masterful works of art were reduced to a list of just three numbers that appear to be able to summarize the characteristic brushstroke and line, what might be called the "style" of their makers? At the heart of the method is a representation of the digital image that is not a simple pointillist collection of pixels, but, rather, one that is more akin to an accumulation or layering of basic linear elements. The latter is a *wavelet representation* and is explained below.

Analysis

While the basic datum of the pixel representation is the pixel, the basic elements of a wavelet representation can be thought of as linear elements of varying scales and orientations. The attribute of orientation is a familiar one and ranges from vertical to horizontal. The notion of scale is perhaps less familiar. It should be thought of as the level of detail at which the analysis is conducted, and it ranges from the finest resolution (as determined by the original digital image) through increasingly coarser representations, or equivalently (using the metaphor of the microscope), from a highest to lowest power of magnification. Finest scale then corresponds to highest magnification, and coarser scales

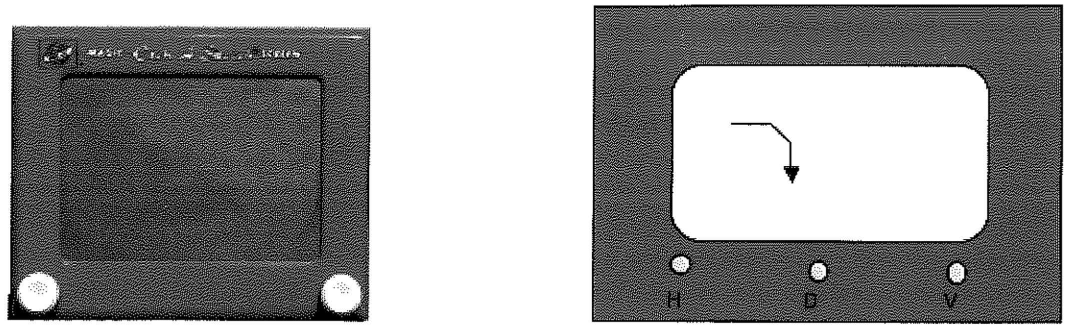

FIGURE 5. On the left is a "standard" Etch-A-Sketch™, where horizontal and vertical moves on the screen are enabled by moving the left and right knobs respectively. At the right is a cartoon of what our wavelet Etch-A-Sketch might look like, this time with three knobs that from left to right control horizontal, diagonal, and vertical movement of the cursor. The three lines on the "screen" are drawn by the independent movement of these three knobs in sequence.

are those determined by lower and lower magnifications. Note that what appears to be a line at a coarse scale might actually be a mish-mash of detailed lines that at a low resolution give an overall linear sense (at some particular orientation).

In discussing the wavelet representation, a useful picture to keep in mind is the well-known "Etch-a-Sketch" children's drawing toy. The standard Etch-a-Sketch has a blank screen and two knobs, one for horizontal navigation and another for vertical navigation. Lines of orientations that are between horizontal and vertical need to be built as varying amounts of horizontal and vertical movements. A wavelet decomposition in essence recovers from the image the amount that the knobs have been turned at each point but with the additional feature that the assumed "wavelet Etch-a-Sketch" has three knobs, permitting vertical, horizontal, and diagonal navigation (Fig. 5). Thus, while the pixel representation implicitly views a digital image as a highly resolved collection of individual light intensities, the wavelet representation can be thought of as a description of the image as a layering, or accumulation, of linear elements of varying length, orientation, and even thickness, at a range of resolutions. The entire process consists of extracting statistics related to this "knob-tuning data" acquired over a range of five scales and then using these numbers as a quantitative signature of the artist.

A picture is worth at least one thousand words (and maybe as many equations), so, using a detail from Bruegel's *Mountain Landscape with Ridge and Valley* (Fig. 1), let us look at some visual representations that illustrate the results of a wavelet analysis (Fig. 6).[13]

As described above, at any particular scale the wavelet decomposition accomplishes the extraction of vertical, horizontal, and diagonal linear elements – i.e., in a given small area (several pixels square) "how much each knob is turned." The results of extracting the vertical elements at the finest scale are shown in Figure 7.[14] Notice how this image is largely composed of vertical or near vertical white lines on a dark background. Generally speaking, regions of white, no matter how small, indicate regions in the original image containing vertical linear elements. This is most apparent in the clear delineation of the vertical lines making up the trunk of the tree in the foreground. While these vertical elements are prominently displayed, the obvious horizontal elements (e.g., the left-right lines indicating much of the leaf work in the tree) are virtually invisible here. Similarly, the horizontal elements represented in Figure 8 emphasize the (largely horizontal) leaf work while neglecting most of the linear elements composing the trunk of the tree.

The "diagonal elements" are also extracted. They are displayed in Figure 9. Admittedly, in this particular detail these are somewhat more difficult to recog-

FIGURE 6. Detail (lower right) of Figure 1, *Mountain Landscape with Ridge and Valley* (c.1552), by Pieter Bruegel the Elder.

FIGURE 7. Vertical elements at the finest level of detail from the lower right of Figure 1, *Mountain Landscape with Ridge and Valley* by Pieter Bruegel the Elder.

nize, but a close inspection of the original image in Figure 6 reveals that the diagonal lines therein provide the most recognizable features in Figure 9.

The analysis part of our process (also called an *algorithm*) subjects the original grayscale digital image to a sequence of such decompositions. Specifically, after extracting vertical, horizontal, and diagonal elements at the finest scale, we repeat the process at the next finest scale. The process of reducing the level of magnification is accomplished by computationally "smoothing" the original image, which is effectively the same as removing the fine details that are only seen at the higher level of magnification. For example, a craggy mountain range when seen at a lower magnification may appear to be less jagged. This effect could be accomplished by removing the smallest zig-zags in the original picture. Thus, the vertical, horizontal, and diagonal information derived at this coarse scale says more about the "sense" of the region of analysis than the particulars. We repeat this process several times, ultimately obtaining the vertical, horizontal, and diagonal information at five scales of analysis. Alternately, this can be thought of as deriving a representation of the original picture as the accumulation of a layering of linear elements that successively provide more and more detail.

SUMMARIZING THE ANALYSIS — From Images To Numbers

We have just described and presented a visual representation of the wavelet decomposition. In essence, the shades of gray at each point in Figures 7-9 are intensity-coded representations of numbers that measure respectively the amount of vertical, horizontal, and diagonal content in a small region about that point. These numbers are extracted at five resolutions or scales. The next step in the process is to understand the distribution of these numbers in a mathematical sense. That is, at any fixed scale (resolution) there is a range of vertical content across the image and we want to summarize this range – i.e., describe the percentage of the image that has a given amount of vertical content. Generally, it can be difficult to describe the full range of behavior, but there are various simple summaries of any collection of numbers that begin to give an idea of their distribution. A typical kind of "summary statistic" would be the *average,* which, roughly speaking, gives an idea of what is the most likely number in the lot. After that we can ask how the collection varies with respect to the average, i.e., the spread, and then the shape of the spread and so on. Thus, do we come to use four standard summaries: the *average, variance, kurtosis*, and *skewness*,[15] thereby representing the overall distribution of each of the vertical, diagonal, and horizontal elements of the image at a given scale by only four numbers. We do this for the three finest levels of detail, thereby acquiring twelve numbers as a partial summary of a given region of the image.

At this point, for purposes of illustration, it is worth drawing an analogy with techniques used in the discipline of "stylometry," which is a mathematical approach to determining literary authorship. The summary statistics that we derive from the wavelet analysis explained above is much like the way in which a stylometric analysis begins to build a numerical signature for literary style based on an author's use of *function words*. Function words carry no meaning without context. Typical examples are prepositions and conjunctions. It turns out that an author's relative predilections for some function words over others as measured by simple word frequencies (e.g., the number of time per thousand words that an author uses the word "however") is a good indicator of individual style and, thus, has helped settle various questions of authorship.[16] We see the "context-free" linear elements derived from a wavelet representation as functioning as a sort of grammatical glue for a picture and their summary statistics as serving a role analogous to the function word usage statistics.

While simple statistical summaries appear to suffice for literature, image analysis seems to need an additional set of numbers that measure the "predictability" of the image. At the finest scale of analysis, this information can be interpreted as providing a mea sure of the continuity of line. More generally, it is simply a measure of the degree to which knowledge of the composition of a picture nearby a location or region is predictive of what is actually present in that region. Mathematically, we derive a uniform

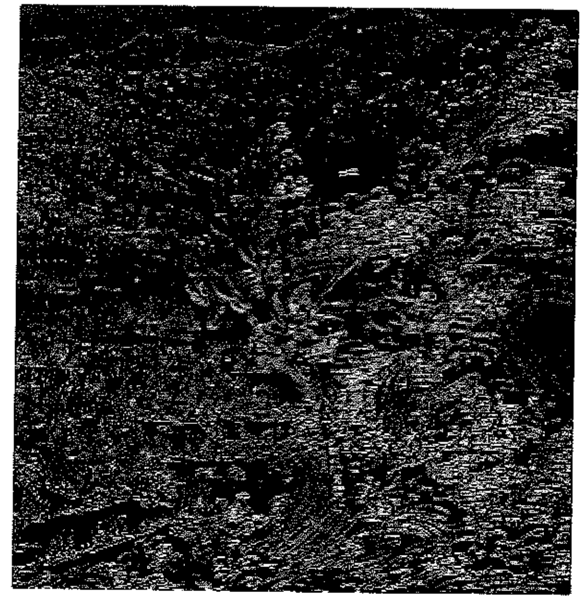

FIGURE 8. Horizontal elements at the finest level of detail from the lower right of *Mountain Landscape with Ridge and Valley*.

way of predicting a given element as a simple combination of nearby elements and then measure the difference between the simple prediction and the truth. We do this for each of the vertical, horizontal, and diagonal elements, which allows us to obtain a distribution of numbers measuring the degree to which this simple prediction is correct. A work that does not have a high degree of local variation will tend to be highly predictable (in this sense) and thus admit a reasonable approximation, measured by many differences being near zero (Fig. 10). The degree of predictability is as much a part of the facture as are the linear predilections measured by the summaries described above. Predictability is also affected by the subject matter of the work. A large patch of clear sky is more predictable than a dense forest. It is for this reason that we compare works with similar subject matter. In the difference between the predicted image and the actual, we find another indication of the artist's style, for, while the large-scale structure of a tree or mountain has a certain basic predictability, it is the manner in which an artist deviates from the structure that defines his or her hand.

As before, a range of numbers are obtained and we again look to summarize the distribution with 4

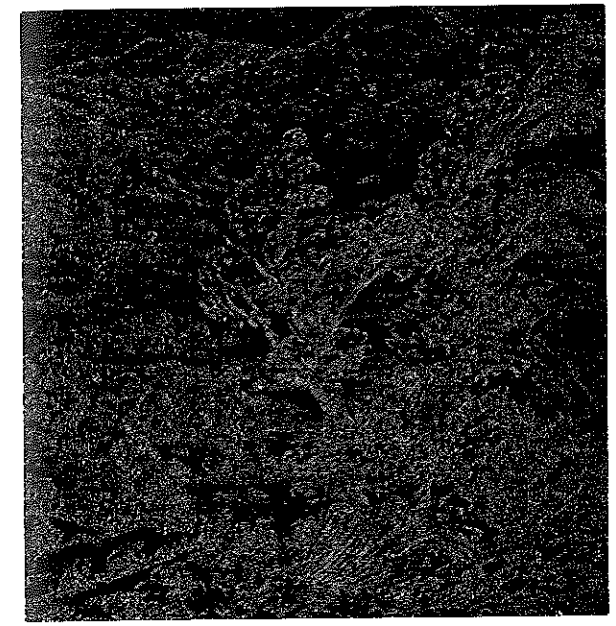

FIGURE 9. Diagonal elements at the finest level of detail from the lower right of *Mountain Landscape with Ridge and Valley*.

numbers (the average, variance, kurtosis, and skewness). Carrying out this procedure for each of the diagonal, horizontal, and vertical elements yields another twelve numbers, so in combination with the twelve statistics of the individual elements (described above), we summarize the wavelet analysis with twenty-four numbers in total. We do this at three levels of detail, so that, ultimately, any region of the original image analyzed in this way is represented by 3 times 24, or 72 numbers.

"The analysis derives a list of 72 numbers from . . . one image. However, . . . we perform the analysis on a collection of sub-images. . . . In the case of the Bruegel analysis, each drawing is reduced to 49 lists of 72 numbers."

THE LAST STEP: From Numbers To A Mathematician's Picture — A Graph

The analysis thus derives a list of 72 numbers from the input of one particular image. However, our technique is such that we perform the analysis described above on a collection of sub-images (details) that provide a tiling of the original picture. In the case of the Bruegel analyses, we independently compute the lists of about 49 contiguous details which together comprise the original. Each block is a square of pixels, 256 pixels on a side. In the case of the facial details of the Perugino, the high resolution of the digital image of the full painting allows us to derive about 100 lists per detail.[17] The challenge of comparing two drawings is thus converted to analysis of these lists of numbers. For example, in the case of the Bruegel analysis, each drawing is now reduced to 49 lists of 72 numbers. As it turns out, we can turn this into a geometry problem in which each picture is now thought of as a "cloud of points" (49 points in the case of Bruegel), and we now need to compute the distance between these clouds as rendered in our familiar three-dimensional space. This is what is shown in Figures 3 and 4. At this stage we have something that we can see and can more easily measure the degree to which there are natural clusters of points, which, in turn, correspond to natural clusters of quantitatively similar works.

CONCLUSION

What we have been describing is a promising computational tool for the digital classification of works of art which, as a *comparative* technique, may be useful for authentication. Our technique looks for consistencies or, as the case may be, inconsistencies in wavelet statistics collected from drawings or paintings (or portions thereof). The technique – based on an admittedly limited universe of Bruegel drawings and a Perugino *Madonna and Child* – seems to work . Of course, there is much work still be done toward understanding and refining the method. To that end, with the help of Walter

"There is much work still to be done . . . We are in the process of acquiring digital images of a group of Rembrandt and Rembrandt-like portraits for the purposes of applying our digital analysis."

Liedtke, Curator of European Paintings at the Metropolitan Museum of Art, and Barbara Bridgers, General Manager for Imaging and Photography at the Metropolitan Museum of Art, we are in the process of acquiring digital images of a group of Rembrandt and Rembrandt-like portraits for the purposes of applying our digital analysis.

Note that this is a comparative technique that enables us to derive some measure of the degree of similarity or dissimilarity among works. Thus, for authentication, we would require a body of work with secure attribution. However, we can also lend some statistical insight to questions of whether or not a corpus of unknown origin has some statistical consistency, and thus, some reasonable chance of being a single person's work.

We look forward to improving these tools through interaction with the art history and conservation communities and hope that this article may generate some new collaborations toward this end. One goal is to achieve a better understanding of the interaction between standard restoration practices and our digital techniques to distinguish the restorer's hand from the artist's. It is also natural to try to incorporate other image data (e.g., underdrawings) into our analysis. We believe that with the growing role of digitization in the arts, that these and other related tools from image processing, used in combination with traditional authentication techniques, will play an important role in the field of art forensics.[18]

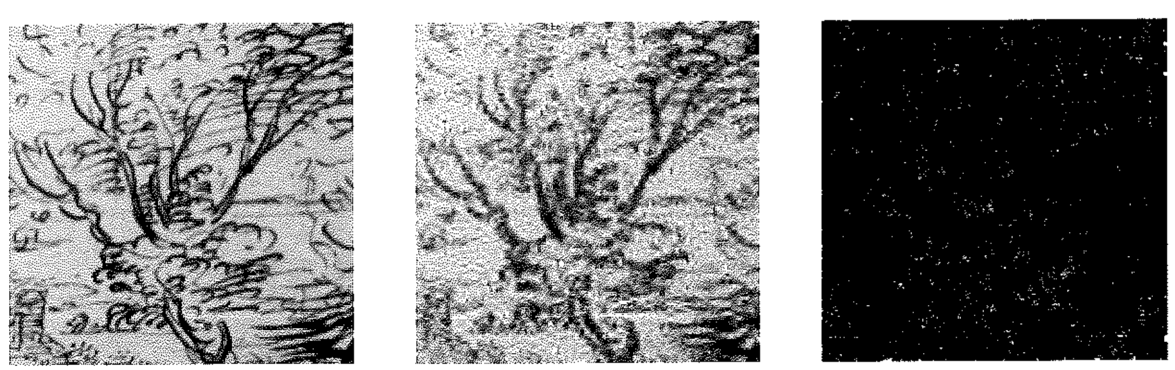

FIGURE 10 A-C. Illustration of predictability. The leftmost figure (10A) is a small detail from *Mountain Landscape with Ridge and Valley*. The center figure (10B) shows the "predicted reconstruction," and 10C shows the difference between leftmost and center. A perfect match would give a pure black square as the difference, so that the lightly speckled difference, image (10C) shows a close match.

ENDNOTES

*Daniel Rockmore, Siwei Lyu, and Hany Farid are professors in the Department of Computer Science and Mathematics, Dartmouth College, Hanover, New Hampshire. The authors welcome comments, which should be addressed to *IFAR Journal* or to D. Rockmore, 6211 Sudikoff Lab, Department of Computer Science, Dartmouth College, Hanover, NH 03755. Tel/Fax: (603) 646-3260/1672; e-mail: rockmore@cs.dartmouth.edu.

1 A good survey of current authentication techniques can be found in *The Expert Versus the Object*, R.D. Spencer, editor. Oxford University Press, 2004.

2 *Pieter Bruegel, Die Zeichnungen*, U. Mielke, editor. Brepols, Turnhout, Belgium, 1996.

3 Our findings confirm the opinion of Hood Museum Curator, Bart Thurber. Nevertheless, the attribution of *Madonna with Child* is the source of a good deal of art historical debate. In particular, Laurence Kanter, Curator of the Robert Lehman Collection European Painting at the Metropolitan Museum of Art, believes that one of the figures may be the work of Perugino's more famous pupil, Raphael.

4 Richard Taylor, "Pollock, Mondrian and Nature: Recent Scientific Investigations." *Chaos and Complexity Letters*, in press. R. Taylor, A.P. Micolich, and D. Jones, "Fractal Analysis of Pollock's Drip Paintings," *Nature*, Vol. 399, 1999, p. 422.

5 A nice review of stylometry can be found in D. I. Holmes and J. Kardos, "Who Was the Author? An Introduction to Stylometry," *Chance*, Vol. 16, no. 2, 2003, pp. 5-8.

6 www.artstor.org

7 The conversion from color to grayscale is accomplished via a standard procedure in which the grayscale value is derived as the intensity of the red, green, blue encoding of the corresponding color pixel. While the result is a loss of color information, the effect is that the measured statistical features and subsequent classification was more likely to be based on the artist's strokes and not on simple color differences. We use an encoding which allows for 256 shades of gray, ranging from zero for the total absence of light to 255 for full presence – i.e., white light.

8 In the Cartesian representation, a list of three numbers indicates three displacements from a reference origin in mutually perpendicular directions, usually denoted as x, y, and z.

9 Our ground truth for Bruegel attribution is U. Mielke, editor, *Pieter Bruegel, Die Zeichnungen*, Brepols, Turnhout, Belgium, 1996. For a recent discussion of some of the issues in Bruegel attributions, see Nadine M. Orenstein, "Followers and Fakers of Pieter Bruegel," *IFAR Journal*, Vol. 6, no. 3, 2003, pp. 12-17.

10 The works used for the Bruegel analysis are (using the catalog numbers from *Pieter Bruegel the Elder: Drawings and Prints, Exhibition Catalogue*, Nadine M. Orenstein, editor. Yale University Press, New Haven and London, 2001) Nos. 3, 4, 5, 6, 9, 11, 13, 20 (attributed to P.B. the Elder) and Nos. 7, 120, 121, 125, 127 (no longer attributed to Pieter Bruegel the Elder).

11 Slides were provided courtesy of Nadine Orenstein, associate curator in the Department of Drawings and Prints at New York's Metropolitan Museum of Art.

12 See discussion, Anne Eisenberg, "Who Really Wielded the Paintbrush?", *The New York Times*, December 23, 2004.

13 A more detailed technical and mathematical discussion of our study can be found in our scientific publications, such as: "Digital techniques for art authentication," S. Lyu, D. Rockmore, and H. Farid, *Proceedings of the National Academy of Sciences*, 87(7):1062-1078, 2004.

14 In order to help clarify the discussion we have heightened the contrast in Figures 8-10.

15 A possibly useful analogy is with the familiar bell-curve, also called the *normal distribution*. A bell-curve is characterized by the position of its center, which is the average of the distribution of the numbers it represents, while the *variance* quantifies the degree to which the bell is either peaked or flattened. The symmetry and rate of this peaking or flattening is futher described the by the *skewness* and *kurtosis*.

16 Perhaps the most famous application of a function word approach to stylometry was the investigation of the disputed essays among the 52 "Federalist Papers," a collection of papers published under the pseudonym "Publicus," but individually written by Alexander Hamilton, James Madison, and John Jay. For details of the story, see F. Mosteller and D. L. Wallace. *Applied Bayesian* and *Classical Inference: The Case of the Federalist Paper.* Springer-Verlag, New York, NY, 1984.

17 From left to right in the painting (see Fig. 2), we cover the six facial details by 189, 171, 189, 54, 81, and 144 blocks, respectively, and derive an equal number of lists of numbers in each case.

18 D. Rockmore has been supported by grant AFOSR F49620-00-1-0280. H. Farid has been supported by an Alfred P. Sloan Fellowship, a National Science Foundation CAREER Award (IIS-99-83806), a departmental National Science Foundation Infrastructure Grant (EIA-98-02068), and under Award No. 2000-DT-CX-K001 from the Office for Domestic Preparedness, U.S. Department of Homeland Security. (Points of view in this document are those of the authors and do not necessarily represent the official position of the U.S. Department of Homeland Security.) We are grateful to George Goldner and Nadine Orenstein of New York City's Metropolitan Museum of Art and Kathy Hart of Dartmouth's Hood Museum of Art for their assistance and cooperation.