



## DIGITAL DETECTIVES REVEAL ART FORGERIES

By Pam Frost Gorder

**A**N EXHAUSTED HANY FARID ANSWERS HIS CELL PHONE FROM A HOTEL IN BOSTON, HIS VOICE HOARSE FROM EIGHT HOURS OF TESTIFYING IN FEDERAL COURT. THE DARTMOUTH

College computer scientist practices a new kind of forensics: he has developed a statistical technique that gauges whether a photograph is computer generated (CG), or if a work of art is a forgery. As a result, he's in high demand.

Reporters want interviews; lawyers—such as Boston prosecutors—want his expert opinion on whether photographic evidence is real or CG (see the sidebar on p. 8); average folks want to know if that painting in the attic is really a Monet. He defers the last group with a form letter.

So far, his computer algorithms have correctly identified five forgeries among 13 artists' drawings and matched some human experts' theories on the origins of a Renaissance oil painting. The results are promising, but Farid and his team will study many more pieces of art before they joke about hiring themselves out as authenticators. "We have a lot to do," he says.

While Farid studies artwork, other scientists are using similar techniques to determine authorship of disputed literary works. Taken together, their findings suggest a surprisingly tangible link between art and science: a painter's every brush stroke and a writer's every word leave behind a unique creative signature—a computer-readable mathematical fingerprint.

### Indelible Fingerprints

Farid had been thinking about the geometry of lines, curves, and shapes drawn at different scales—that is, the mathematics of images—long before he walked into a special exhibition at New York City's Metropolitan Museum of Art in 2001. The exhibition featured drawings by Flemish artist Pieter Bruegel the Elder, as well as some high-quality fake Bruegels. In those landscapes' fine lines and shading, Farid saw visual elements that would lend themselves to a mathematical technique called *wavelet analysis*.

Wavelets can break down a picture into vertical, horizontal, and diagonal elements on large and small scales. Statistical algorithms can then detect a pattern in those elements—an image's unique signature.

Farid knew from his own research that a real photograph has a different signature than a CG one. Objects within an image have different signatures, too; the sky on a landscape registers differently than grass or a house. As he stood in the museum, he wondered if a fake Bruegel would register differently than the real thing.

According to the museum, Bruegel was one of the 16th century's most influential artists, not just because of his skill with a brush or pen, but because he was able to infuse landscapes with great feeling. After growing up in the lowlands of northern Europe, a trip through the Italian Alps made such an impression on him that he drew detailed mountain scenes long after returning home.

Forgers tried to imitate Bruegel's style with their own Alpine landscapes, but did they succeed? To find out, Farid, Dartmouth colleague Daniel Rockmore, and student Siwei Lyu looked for patterns that would distinguish one artist from another. They scanned 35-mm photographs of the drawings and deconstructed the images with wavelets. Then, starting with the algorithms Farid had developed to tell real images from CG, they plugged new algorithms into a small Beowulf computer cluster.

Next, the scientists used another mathematical technique called multidimensional scaling (MDS) to shrink the 72 model variables—72 dimensions of data—down to three. Each drawing became a point in 3D space, with similarly styled drawings appearing closer together. Viewed this way, the eight genuine Bruegels clustered together whereas the five fakes stood out, scattered around the periphery.

Fresh from this discovery, and "as much for fun as anything else," Farid and his team decided to try their algorithms on a different kind of artwork, the 15th century oil painting *Madonna and Child* from the Hood Museum of Art at Dartmouth. Historians believe that the Italian artist Pietro Perugino collaborated with some of his students on the painting, but they don't agree on who painted which parts. So, Farid's attempt to solve the mystery was "a bit of

---

## EDITOR IN CHIEF

**Norman Chonacky**  
cise-editor@aip.org

---

## ASSOCIATE EDITORS IN CHIEF

Denis Donnelly, Siena College  
donnelly@siena.edu

Douglass E. Post, Los Alamos Nat'l Lab.  
post@lanl.gov

John Rundle, Univ. of California at Davis  
rundle@physics.ucdavis.edu

Francis Sullivan, IDA Ctr. for Computing Sciences  
fran@super.org

---

## EDITORIAL BOARD MEMBERS

Klaus-Jürgen Bathe, Mass. Inst. of Technology, kjb@mit.edu

Antony Beris, Univ. of Delaware, beris@che.udel.edu

Michael W. Berry, Univ. of Tennessee, berry@cs.utk.edu

John Blondin, North Carolina State Univ.,  
john\_blonidin@ncsu.edu

Michael J. Creutz, Brookhaven Nat'l Lab., creutz@bnl.gov

George Cybenko, Dartmouth College, gvc@dartmouth.edu

Jack Dongarra, Univ. of Tennessee, dongarra@cs.utk.edu

Rudolf Eigenmann, Purdue Univ.,  
eigenman@ecn.purdue.edu

David Eisenbud, Mathematical Sciences Research Inst.,  
de@msri.org

William J. Feiereisen, Los Alamos Nat'l Lab,  
bill@feiereisen.net

Geoffrey Fox, Indiana State Univ., gcf@grids.ucs.indiana.edu

Sharon Glotzer, Univ. of Michigan, sglotzer@umich.edu

Anthony C. Hearn, RAND, hearn@rand.org

Charles J. Holland, Office of the Defense Dept.,  
charles.holland@osd.mil

M.Y. Hussaini, Florida State Univ., myh@cse.fsu.edu

David P. Landau, Univ. of Georgia,  
dlandau@hal.physast.uga.edu

B. Vincent McKoy, California Inst. of Technology,  
mckoy@its.caltech.edu

Jill P. Mesirov, Whitehead/MIT Ctr. for Genome Research,  
mesirov@genome.wi.mit.edu

Charles Peskin, Courant Inst. of Mathematical Sciences,  
peskin@cims.nyu.edu

Constantine Polychronopoulos, Univ. of Illinois,  
cdp@csrds.uiuc.edu

William H. Press, Los Alamos Nat'l Lab., wpress@lanl.gov

John Rice, Purdue Univ., jrr@cs.purdue.edu

Ahmed Sameh, Purdue Univ., sameh@cs.purdue.edu

Henrik Schmidt, MIT, henrik@keel.mit.edu

a parlor trick because nobody knows for sure what the right answer is," he says.

The scientists analyzed the face of each figure in the painting as they had the Bruegel drawings, hoping that minute differences in the contributors' brush strokes would stand out. The computer's finding—that Perugino painted the Madonna and two of the saints in the image while his students filled in the rest—meshes with some experts' opinions. Those results recently appeared in the *Proceedings of the National Academy of Sciences* (vol. 101, no. 49, 2004, pp. 17006–17010).

## Paint Cracks

Like Farid, other scientists are using statistics to learn more about art. At the Fitzwilliam Museum at the University of Cambridge, conservator Spike Bucklow is studying the cracks that naturally form in oil paintings over time. Since 1998, he's been developing models to tie the crack patterns, known as *craquelure*, to a particular craft tradition.

Some fissures trace out delicate grid lines, whereas others follow a jagged path like lightning. The pattern depends on an artist's materials: the paint chemistry, application method, and canvas. Because artists who lived in a particular country at a particular time tended to use similar materials, *craquelure* can help confirm where and when a work was painted.

Bucklow examined 600 paintings from four distinct categories: Italian, Flemish, Dutch, and French, from the 14th to the 18th centuries. He modeled the crack patterns and then reduced his data, as Farid did, with MDS. In the journal *Computers and the Humanities*, he reported that paintings from each region did indeed possess a characteristic *craquelure* (vol. 31, 1998, pp. 503–521).

Some tricky artists fooled the computer, though: an Italian painter who used northern European techniques registered as Flemish, and a Dutch painter who used Parisian art supplies registered as "almost more French than the French." Overall, Bucklow's statistical method matched paintings to the right country and century 97 percent of the time. He's continuing the work, and says he can now identify some paintings by city and by decade.

He feels that the strength of Farid's technique lies in his drawing analysis, where the narrow pen strokes are more two-dimensional, like *craquelure*. To Bucklow, the texture of oil paint and the addition of color make *Madonna and Child* a more complicated, and less reliable, test case.

Farid definitely wants to look at more paintings, ones that human authenticators can all agree on. Then, he says, he can think about tackling more controversial works of art.

## Computational Chemistry and Biology Opportunities at D. E. Shaw Research and Development

### Battling Bards

In literary circles, few controversies elicit more venom than debates over William Shakespeare. The Bard's true identity, and whether all his plays were actually written by the same person, are both contentious points.

Statistical research at Harvard Medical School is contributing to the debate. There, physician Ary Goldberger and his team have taken the algorithms they use to diagnose abnormal heart rhythms and adapted them to study word patterns in Shakespeare and other literary works.

They first realized that their algorithms could be useful outside of medicine when they began translating their heart readings into binary series. The heartbeats emerged on a computer in a string of short and long sequences, which Goldberger's colleagues Albert Yang and C.K. Peng began to think of as words in a sentence. To Goldberger, assessing a person's heart health with this technique became like reading the patient's biography.

If the Harvard scientists could detect patterns in the pseudo-text of someone's heartbeat, they wondered, could they detect patterns in real text, too? Using the algorithms that Peng and Yang developed, they mapped the frequency of words in plays by Shakespeare and some of his contemporaries into graphical tree structures that showed how the writers' styles related to each other and how they evolved over time.

In a paper in the journal *Physica A* (vol. 329, 2003, pp. 473–483), they concluded that Shakespeare's tree trunk started out close to that of another writer, Christopher Marlowe, but then evolved in a different direction, a finding that agrees with some views that Marlowe influenced a young Shakespeare. One of the plays controversially credited to Shakespeare appeared solidly on a branch of Marlowe's tree, indicating that he was the major author.

"This type of analysis is never proof in the sense of a smoking gun," Goldberger says, but because he and his team validated their findings with several different literary databases, he feels their statistics are robust.

Peng says he thinks their technique could even uncover patterns in large research databases, where a simple keyword search might not pick up on more complex relationships among data. Before that can happen, though, someone would have to develop a scheme to parallelize the algorithms for more efficient computation, possibly on a supercomputer.

The Harvard team is now working to apply the algorithms to biomedical databases, images, and artwork. They've posted their heartbeat data and related algorithms online at

Extraordinarily gifted computational chemists, biologists, and other computational scientists are sought to join a rapidly growing New York-based research group that is pursuing an ambitious, long-term strategy aimed at fundamentally transforming the process of drug discovery.

Candidates should have world-class credentials in computational chemistry, biology, or physics, or in a relevant area of computer science or applied mathematics, and must have unusually strong research skills. Relevant areas of experience might include protein structure prediction, the computation of protein-ligand binding affinities, the study of biologically important systems using molecular dynamics and/or Monte Carlo simulation, and the application of statistical mechanics to biomolecular systems—but specific knowledge of any of these areas is less critical than exceptional intellectual ability and a demonstrated track record of achievement. Current areas of interest within the group include molecular dynamics simulation of functionally significant globular and membrane proteins, the prediction of protein structures and binding free energies, structure- and ligand-based drug design, characterization of protein-protein, protein-nucleic acid and protein-lipid interactions, and the development of algorithms for biomolecular simulations.

This research effort is being financed by the D. E. Shaw group, an investment and technology development firm with more than \$10 billion in aggregate capital. The project was initiated by the firm's founder, Dr. David E. Shaw, and operates under his direct scientific leadership.

We are eager to add both senior- and junior-level members to our world-class team, and are prepared to offer above-market compensation to candidates of truly exceptional ability. Please send your CV (including list of publications, thesis topic, and advisor, if applicable) to [computing@desrad.deshaw.com](mailto:computing@desrad.deshaw.com).

D. E. Shaw Research and Development, L.L.C. does not discriminate in employment matters on the basis of race, color, religion, gender, national origin, age, military service eligibility, veteran status, sexual orientation, marital status, disability, or any other protected class.

The logo for D E Shaw & Co features the company name in a serif font, with a stylized horizontal line above the 'E' and 'S' that tapers at both ends.

## EDITORIAL OFFICE

COMPUTING in SCIENCE & ENGINEERING  
10662 Los Vaqueros Circle, PO Box 3014  
Los Alamitos, CA 90720  
phone +1 714 821 8380; fax +1 714 821 4010;  
www.computer.org/cise/

## DEPARTMENT EDITORS

**Book & Web Reviews:** Bruce Boghosian, Tufts Univ., bruce.boghosian@tufts.edu

**Computing Prescriptions:** Isabel Beichl, Nat'l Inst. of Standards and Tech., isabel.beichl@nist.gov, and Julian Noble, Univ. of Virginia, jvn@virginia.edu

**Computer Simulations:** Dietrich Stauffer, Univ. of Köhn, stauffer@thp.uni-koeln.de

**Education:** Denis Donnelly, Siena College, donnelly@siena.edu

**Scientific Programming:** Paul Dubois, Lawrence Livermore Nat'l Labs, dubois1@llnl.gov, and George K. Thiruvathukal, gkt@nimkathana.com

**Technology News & Reviews:** Norman Chonacky, Columbia Univ., norman.chonacky@yale.edu

**Visualization Corner:** Jim X. Chen, George Mason Univ., jchen@cs.gmu.edu, and R. Bowen Loftin, Old Dominion Univ., bloftin@odu.edu

**Your Homework Assignment:** Dianne P. O'Leary, Univ. of Maryland, oleary@cs.umd.edu

## STAFF

**Senior Editor:** Jenny Ferrero, jferrero@computer.org

**Group Managing Editor:** Gene Smarte

**Staff Editors:** Kathy Clark-Fisher, Rebecca L. Deuel, and Steve Woods

**Production Editor:** Monette Velasco

**Magazine Assistant:** Hazel Kosky, cise@computer.org

**Technical Illustrations:** Alex Torres

**Publisher:** Angela Burgess, aburgess@computer.org

**Assistant Publisher:** Dick Price

**Advertising Coordinator:** Marian Anderson

**Marketing Manager:** Georgann Carter

**Business Development Manager:** Sandra Brown

## AIP STAFF

Jeff Bebee, Circulation Director, jbebee@aip.org

Charles Day, Editorial Liaison, cday@aip.org

## IEEE ANTENNAS AND PROPAGATION SOCIETY LIAISON

Don Wilton, Univ. of Houston, wilton@uh.edu

## IEEE SIGNAL PROCESSING SOCIETY LIAISON

Elias S. Manolakos, Northeastern Univ., elias@neu.edu

## CS MAGAZINE OPERATIONS COMMITTEE

Bill Schilit (chair), Jean Bacon, Pradip Bose, Doris L. Carver, Norman Chonacky, George Cybenko, John C. Dill, Frank E. Ferrante, Robert E. Filman, Forouzan Golshani, David Alan Grier, Rajesh Gupta, Warren Harrison, James Hendler, M. Satyanarayanan

## CS PUBLICATIONS BOARD

Michael R. Williams (chair), Michael R. Blaha, Roger U. Fujii, Sorel Reisman, Jon Rokne, Bill N. Schilit, Nigel Shadbolt, Linda Shafer, Steven L. Tanimoto, Anand Tripathi



IEEE Antennas &  
Propagation Society

IEEE  
Signal Processing Society

## Real vs. Computer Generated

Many Farid wants to study more artwork in the future, but right now he's busy using his statistical technique to distinguish real photographs from computer-generated (CG) ones. He found a new purpose for his work in 2002, when the Supreme Court ruled that CG, or "virtual" child pornography was protected by the First Amendment. Since then, defendants have claimed that the digital pornographic images they're being prosecuted for are actually virtual, so prosecutors have had to prove that the images are real. The differences can be hard to see, and Farid says he thinks his technique might be the only one that can do so in a quantifiable, scientific way.

So far, he's provided expert testimony for two cases, and both times the children's images turned out to be real photographs. He's not surprised. "The fact is, people who are doing CG are not making child pornography," he says.

He endures long hours on the stand because courts have to decide whether his technology is admissible, and that's okay with him. DNA fingerprinting went through similar rigors before it became accepted evidence, so why shouldn't his image analysis? "It's a relatively new science, which I'll be the first to admit," he says. "But at some point, something like this has to get into the courts because right now it's very hard to prosecute child pornography."

In the meantime, he just received a grant to integrate his technology into FBI crime labs, so wavelet analysis could soon find even more avenues into the courtroom.

www.physionet.org, an open-source site sponsored by the National Center for Research Resources at the US National Institutes of Health.

Goldberger sees a natural connection between a heartbeat and creativity patterns that artists and authors express. "Works of art are physiologic signals, to the extent that they are produced by the human brain," he says. "Musical compositions are the same, and so are works of art and literature."

With further development, these statistical methods could join the art authenticator's standard toolkit alongside physical techniques such as x-ray radiography and chemical analysis. Technology could help resolve all-too-human controversies in art and authorship.

The computer has confirmed one thing: Shakespeare might have been influenced by Marlowe, and Bruegel's heartfelt landscapes might have inspired many imitators, but ultimately, nothing could hide either master's signature style.

Pam Frost Gorder is a freelance science writer based in Columbus, Ohio.