The Van Gogh Project: Art Meets Mathematics in Ongoing International Study

By Michelle Sipics

In a digital world, style is mathematical.

So said Dan Rockmore, a professor of mathematics at Dartmouth College, in the I.E. Block Community Lecture at last year's SIAM Annual Meeting. According to Rockmore, "We are, in part, our 'numbers'—and if you leave a trail of numbers, it makes sense to try to look at them."

Rockmore's numbers represent aspects of style. They're a quantification of a painter's brush strokes, a dancer's movements, a writer's word choices, a composer's notes. A mathematician, says Rockmore, can analyze a work of art and transform it into a collection of numbers, which can then be used as the basis for further conclusions about the work.

In his own work, Rockmore has analyzed drawings by the 16th-century Flemish artist Pieter Bruegel (see "Can Mathematical Tools Illuminate Artistic Style," *SIAM*

Detail from the painting "Willows at Sunset" showing the use of a new algorithm to isolate brushstrokes. The painting's interest to the Van Gogh Project lies in part in the challenge of dating it, given its stylistic elements from two periods in the artist's career. Not universally acknowledged to be a genuine Van Gogh, "Willows" has also been the focus of the project's attribution efforts. Courtesy of the Kröller-Müller Museum and the James Z. Wang Research Group at Penn State.

News, March 2005; http://www.siam.org/news/news.php?id=34). He discussed that work in the Block lecture, along with his participation in an ongoing collaborative effort to analyze works—paintings this time—of one of the most famous artists of all time: Vincent Van Gogh.

Analyzing the Dutch Master

The Van Gogh Project, now in its third phase, is the brainchild of Rick Johnson, a professor of electrical and computer engineering at Cornell University and, as of 2007, an adjunct research fellow of the Van Gogh Museum in Amsterdam. (His 1977 PhD, from Stanford University, was in electrical engineering and included a minor in art history, with a specialty in Dutch art.) He spent much of a 2005 sabbatical looking for a way to combine his dual interests.

During that sabbatical, Johnson encountered the work of three researchers who were using statistics for image analysis: Ingrid Daubechies of Princeton, James Wang of Penn State, and Eric Postma (then of Maastricht University in the Netherlands, now at Tilburg University). Postma was studying paintings by Van Gogh.

Johnson struck a deal with the Van Gogh Museum, securing near-unprecedented access to high-resolution scans of more than 100 paintings that he and his collaborators would analyze for authenticity. This was the birth of the Van Gogh Project, a collaborative undertaking by the museum and its curators and conservators, as well as mathematicians and computer scientists. The goal of the first phase: to generate a mathematical and computational characterization of the Dutch artist's technique, including brushwork, composition, and color choices.

Given Johnson's art history specialization, Van Gogh was a natural choice. But Johnson also points to an element of practicality in the selection of the Dutch master.

"The holdup in doing this kind of research is that you can't get your hands on really good data," says Johnson. "If you pick someone whose paintings are [spread out] everywhere, you have to go everywhere and beg. So the Van Gogh Museum is perfect for this."

Once the museum was on board, Johnson had to convince mathematical researchers to participate in the project—and to view it as a genuine collaboration, a sharing of goals and expertise among themselves and the art historians. He recalls his pitch to prospective participants: "If I can get you this data, you've got to do a day-long workshop, no equations, explaining what this can do."

Daubechies, Wang, and Postma were among those recruited for the first phase (2007) of the Van Gogh Project. (Postma took over from Johnson as head of the project in 2008.) The researchers worked in teams; two of the mathematical teams were based at Penn State and Princeton. The first phase of the project concluded with a workshop at the museum in Amsterdam. There, the collaborative spirit called for by Johnson was critical: The mathematical researchers described the tools and concepts used in their research to the art historians, explaining how things like wavelets and Fourier transforms could contribute to research in art authentication. (Ingrid Daubechies, for example, gave a presentation called "Introduction to Wavelets," using Van Gogh paintings to provide contextual examples.) The art historians, for their part, explained more about the process they use in authentication and classification.

Flattery, Forgery, or Original?

As the Block lecturer, Rockmore faced much the same challenge. For the non-mathematicians in the audience, he proceeded by analogy, likening the use of wavelet statistics in image analysis to a child drawing with an Etch-A-Sketch. The child creates a drawing by twisting a pair

of knobs that control the horizontal and vertical movements of a stylus positioned behind the toy's screen—eventually resulting in an image made from one continuous line.

Imagine, Rockmore said, that you take a tiny patch of a finished Etch-A-Sketch image and ask how much horizontal, vertical, and diagonal sketching was done, on average, in that patch. In essence, this is how wavelet decomposition is used to obtain a multidimensional summary of an image.

In their work on Bruegel's drawings, Rockmore, Dartmouth colleague Hany Farid, and Farid's student Siwei Lyu (who is now at the University of Albany) used such summaries to test the authenticity of some of the artist's drawings. As reported in a 2004 paper in the *Proceedings of the National Academy of Sciences*, the researchers studied eight Bruegel drawings, along with five known forgeries or imitations. They divided digital versions of each into non-overlapping regions, creating 64 subimages. Wavelet-like decompositions then yielded a feature vector of 72 coefficient and error statistics for each subimage.

In the I.E. Block Community Lecture at the 2008 SIAM Annual Meeting, Dan Rockmore made a convincing case for the effectiveness of mathematics in analyzing works of art. Wavelet analysis in particular has played a central role in the ongoing Van Gogh Project. From Rockmore's Block lecture.

Each drawing, Rockmore explained, now corresponded to a set of points in a 72-dimensional space; by computing the Hausdorff distance between all possible pairs of images, and subjecting the resulting distance matrix to multidimensional scaling, the researchers were able to plot a point for each drawing in a lower-dimensional space. As a result of the differences between the drawings—reflected in the feature vectors—authentic drawings were grouped together in the graph, while the forgeries lay outside.

Related research was performed during the first phase of the Van Gogh Project, which culminated with a workshop titled "The First International Workshop on Image Processing for Artist Identification: Brushwork in the Paintings of Vincent Van Gogh." The full proceedings are available at the project's Web site, www.digitalpaintinganalysis.org.

Same Artist, New Challenges

Rockmore, who also participated in the second phase of the Van Gogh Project, looked ahead in the Block lecture to the October 2008 workshop in Amsterdam. He highlighted new challenges posed for that part of the project: dating, attribution, and distinguishing feature extraction.

The dating of paintings might seem to be a question more for chemical analysis than for statistics. But each of the challenges, Rockmore says, calls for a multimodal study.

"I don't think mathematics would be *the* answer to any question in any of these analyses," he says. "The solution to the differential equations isn't going to tell you that this painting was made on January 2nd, 1842; you're going to get some big window. With any of these questions, it's like you're building a case."

That said, he continues, a strong analogy for the use of mathematics to date a painting can be found in literary stylometry. In the Block lecture, he cited the example of Wincenty Lutoslawski's efforts, in the late 19th century, to date Plato's *Dialogues* via analysis of the texts.

Lutoslawski wanted to compute some sort of statistic about the way Plato wrote. With data for each of Plato's works, he would create a plot over time. Assuming that such a plot would result in some kind of smooth curve through the points, he would compute that same statistic for an unknown work and place it where it "fit" along one curve—thus determining an approximate date for it.
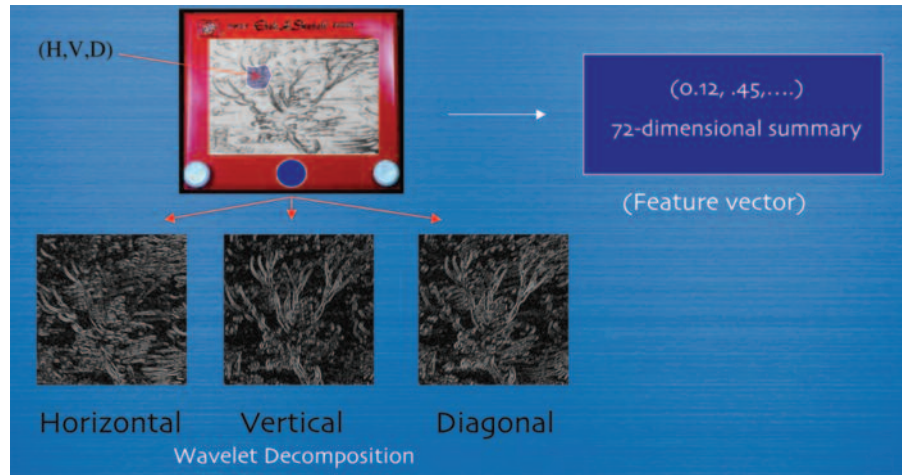
"That was his original motivation for investigating the possibilities of what we could call feature extraction from text," Rockmore explains, "not just for categorizing style, but also for investigating the evolution of style and trying to infer when things were written."

Participants in the Van Gogh Project have used similar techniques to distinguish Van Gogh's paintings by the period in which they were created. Art historians had identified several general differences between the paintings from the artist's Paris period and his later paintings, done in southern France. Smaller brushstrokes are more common in the paintings from Paris, for example; those from the south of France include more contour lines and greater color saturation. In an ongoing effort, the researchers are working to date three "unresolved" Van Goghs—"Still Life: Potatoes in a Yellow Dish," "Willows at Sunset," and "Crab on its Back"—that have characteristics of both periods.

Not all experts accept "Willows at Sunset" as a genuine Van Gogh, which leads to another ongoing project challenge: attribution. In one of their first attacks on this question, the researchers used two training sets, one of paintings known to be by Van Gogh and the other of work by his contemporaries; they then attempted to associate a single test painting with one group or the other based on distinguishing numerical features (brushwork anaylsis was the major factor here), thus making a case for attribution. Others have approached the problem of attribution by analyzing the texture and brushstrokes of as many Van Goghs as possible, and then determining which paintings are most like the others in terms of each individual feature. (Jia Li and James Wang of Penn State, for example, found "Willows at Sunset" to be among the works with texture least like that of the paintings in the Van Gogh training set.)

Proof—of Concept

All of this, Rockmore is quick to point out, is simply evidence, and not proof. Still, while the researchers in the Van Gogh Project may not be able to definitively prove authenticity or date of creation with their analyses, the first year's workshop did accomplish something in the nature

of a proof: a proof of concept. The researchers were able to demonstrate that statistical techniques could, in fact, offer valid analysis of and information about an artist's working style.

The results of the first workshop also had a secondary effect: an establishment of trust between the applied mathematicians and the art historians. It's very rare for museums to give outside individuals access to their artwork, or even scans of their artwork, for fear of unauthorized reproductions. Even for Johnson's project, the Van Gogh Museum limited the researchers to scans of the paintings with half of each image converted to grayscale, so as to minimize the possibility of the high-resolution scans being used to generate copies.

The collaboration among art historians and mathematicians at the first workshop produced a solid foundation for the second, and now a third, year of research.

Looking Toward the Future

The third workshop, planned for the spring of 2010 at the Museum of Modern Art in New York, will be run by Princeton postdoc Shannon Hughes, a former PhD student of Daubechies. Hughes, who will become an assistant professor at the University of Colorado at Boulder this summer, points to a new focus for the researchers.

"While interesting to us, and very promising, these results are not necessarily so compelling to an art expert. An art expert wants to better understand a specific work; to understand not just that we can distinguish between different artists, for example, but what the distinguishing characteristics are. In short, an art expert is interested in making more subtle and nuanced observations than the binary yes-or-no answers that a classifier spits out."

"In the first two workshops, Rick worked hard to translate the art historical issues that he felt image analysis might be able to address into problems that we on the technical side could latch onto," she says. "He asked the art historians to suggest issues they were interested in, and framed each of these issues for us in the format of a standard machine learning problem."

As a result, Hughes says, they've obtained some fascinating results, such as an 85% classification accuracy in distinguishing Van Goghs from paintings by other artists. But there's one problem.

"While interesting to us, and very promising, these results are not necessarily so compelling to an art expert," Hughes explains. "An art expert wants to better understand a specific work; to understand not just that we can distinguish between different artists, for example, but what the distinguishing characteristics are. In short, an art expert is interested in making more subtle and nuanced observations than the binary yes-or-no answers that a classifier spits out."

Having developed some reasonably successful tools, she continues, "We want to try to deploy some of them in real situations and see what they can do." This will be the goal of the 2010 workshop, she says; "each team will consist of both an art historical and a technical component, [and] the two sides will have to work together to do something of art historical interest."

It's the desire to create alliances and shared advances between fields that keeps Johnson involved. The project has opened several doors for him, he says, but in the end, what's exciting is the interdisciplinary nature of the project—and the possibility of creating a new field of study.

"For me," he says, "it's all about trying to make the area explode."

Michelle Sipics is a contributing editor at SIAM News.