

## Dartmouth College Dartmouth Digital Commons

---

Open Dartmouth: Faculty Open Access Articles

---

1-26-2010


# Quantification of Artistic Style through Sparse Coding Analysis in the Drawings of Pieter Bruegel the Elder

James M. Hughes  
*Dartmouth College*

Daniel J. Graham  
*Dartmouth College*

Daniel N. Rockmore  
*Dartmouth College*

Follow this and additional works at: <https://digitalcommons.dartmouth.edu/facoa>

 Part of the [Arts and Humanities Commons](#), [Computer Sciences Commons](#), [Mathematics Commons](#), and the [Statistics and Probability Commons](#)

---

### Recommended Citation

Hughes, James M.; Graham, Daniel J.; and Rockmore, Daniel N., "Quantification of Artistic Style through Sparse Coding Analysis in the Drawings of Pieter Bruegel the Elder" (2010). *Open Dartmouth: Faculty Open Access Articles*. 1486.  
<https://digitalcommons.dartmouth.edu/facoa/1486>

This Article is brought to you for free and open access by Dartmouth Digital Commons. It has been accepted for inclusion in Open Dartmouth: Faculty Open Access Articles by an authorized administrator of Dartmouth Digital Commons. For more information, please contact [dartmouthdigitalcommons@groups.dartmouth.edu](mailto:dartmouthdigitalcommons@groups.dartmouth.edu).

# Quantification of artistic style through sparse coding analysis in the drawings of Pieter Bruegel the Elder

James M. Hughes<sup>a</sup>, Daniel J. Graham<sup>b</sup>, and Daniel N. Rockmore<sup>a,b,c,1</sup>

Departments of <sup>a</sup>Computer Science and <sup>b</sup>Mathematics, Dartmouth College, Hanover, NH 03755; and <sup>c</sup>Santa Fe Institute, Santa Fe, NM 87501

Edited by David Mumford, Brown University, Providence, RI, and approved November 30, 2009 (received for review September 14, 2009)

Recently, statistical techniques have been used to assist art historians in the analysis of works of art. We present a novel technique for the quantification of artistic style that utilizes a sparse coding model. Originally developed in vision research, sparse coding models can be trained to represent any image space by maximizing the kurtosis of a representation of an arbitrarily selected image from that space. We apply such an analysis to successfully distinguish a set of authentic drawings by Pieter Bruegel the Elder from another set of well-known Bruegel imitations. We show that our approach, which involves a direct comparison based on a single relevant statistic, offers a natural and potentially more germane alternative to wavelet-based classification techniques that rely on more complicated statistical frameworks. Specifically, we show that our model provides a method capable of discriminating between authentic and imitation Bruegel drawings that numerically outperforms well-known existing approaches. Finally, we discuss the applications and constraints of our technique.

art analysis | art authentication | image classification | machine learning | stylometry

The use of mathematical and statistical techniques for the analysis of artwork generally goes by the name “stylometry.” Although such qualitative techniques have a fairly long history in literary analysis (see, e.g., ref. 1), their use in paintings and drawings is much more recent. Although this work is still in its relative infancy, statistical methods have shown potential for augmenting traditional approaches to the analysis of visual art by providing new, objective, quantifiable measures that assess artistic style (2–5), as well as other perceptual dimensions (6–9). Recent studies have shown that mathematical analyses can produce results in line with accepted art historical findings (2, 10).

The statistical approaches that can be applied to the analysis of artistic style are varied, as are the potential applications of these approaches. Wavelet-based techniques are often used [e.g., (2)], as are fractals (3), as well as multiresolution hidden Markov methods (11). In this paper, we bring instead the adaptive technique of sparse coding to bear on the problem. Although originally developed for vision research (12), we show that the principle of sparse coding (finding a set of basis functions that is well-adapted for the representation of a given class of images) is useful for accomplishing an image classification task important in the analysis of art. In particular, we show that a sparse coding model is appropriate for distinguishing the styles of different artists. This kind of discriminatory ability could be used to provide statistical evidence for, or against, a particular attribution, a task which is usually known as “authentication.”

In this paper, we consider the application of sparse coding to a particular authentication task, looking at a problem that has already been attacked by statistical techniques (2, 10): distinguishing a set of secure drawings by the great Flemish artist Pieter Bruegel the Elder (1525–1569) from a set of imitation Bruegels, each of whose attribution is generally accepted among art historians. The drawings in the group of imitations were long thought to be by Bruegel (13), so that their comparison to secure Bruegels is especially interesting.

The sparse coding model attempts to create the sparsest possible representation of a given image (or set of images). Thus, a useful statistic for the attribution task is to compare the kurtosis of the representations of the authentic and imitation Bruegels in order to determine their similarity to a control set of authentic Bruegel drawings. Fig. 1 shows the steps involved in our analysis.

We find that a sparse coding approach successfully distinguishes the secure Bruegel drawings from the imitations. In addition, we compare our method to two other approaches: (i) the technique in refs. 2 and 10 that uses wavelets (quadrature mirror filters), and (ii) an approach similar to ref. 2 in which the wavelets are replaced by *curvelets*. Curvelets form a tight frame for the image space and are well adapted for generating sparse representations of curves, thus making them potentially appropriate for representing art images (14, 15). We find that sparse coding is superior (from a statistical point of view) for the task of discriminating imitations from secure Bruegels. These findings suggest that sparse coding could be of great use in visual stylometry.

## Sparse Coding Model

The sparse coding model attempts to describe a space of images by training an overcomplete set of not necessarily orthogonal basis functions that describe the space optimally according to a sparseness constraint (12). Although sparse coding is quite similar to independent component analysis (ICA), it differs in that in sparse coding the kurtosis of the resultant output distribution is maximized, whereas in ICA, the statistical independence of the outputs is maximized. However, if the sparseness constraint applied in a sparse coding model is the logarithm of a sparse probability density function, then the sparseness measure is essentially equivalent to entropy in an information-theoretic approach to ICA (16).

The use of sparse coding in the 2D image domain was inspired by work in vision science related to natural scenes. Olshausen and Field showed that training a sparse coding model to represent the space of natural images resulted in basis functions that were similar to receptive fields in primate visual cortex (12). It is generally believed that these functions capture the localized orientation and spatial frequency information that exists in natural scenes (17).

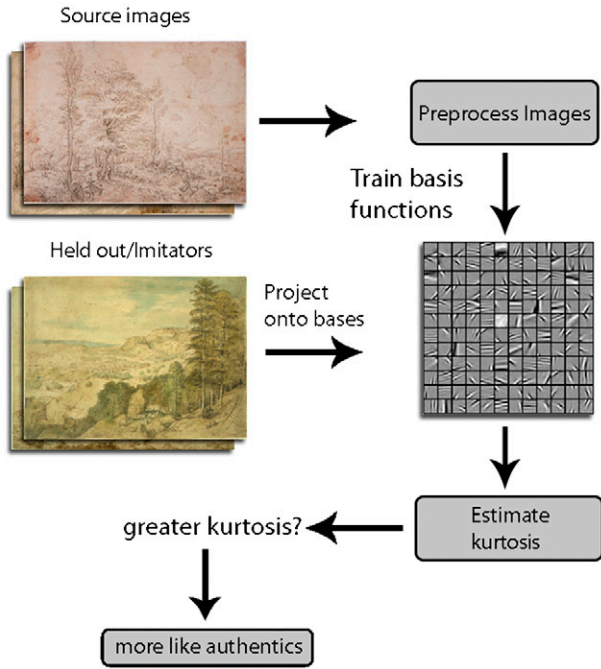
The success of sparse coding in the analysis of natural scenes indicates that it could be useful for modeling features in drawings and other 2D media. Indeed, it has been suggested that the sparseness of the statistical structure of works of art, which can only be detected via higher-order statistics, may contribute to the perception of similarity (6, 7, 18). The sparse coding model, working with small local patches of images, learns functions that could capture the properties of a particular artist's style, to the extent that these properties are perceptible.

Author contributions: J.M.H., D.J.G., and D.N.R. designed research; J.M.H. and D.J.G. performed research; J.M.H., D.J.G., and D.N.R. contributed new reagents/analytic tools; J.M.H., D.J.G., and D.N.R. analyzed data; J.M.H., D.J.G., and D.N.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

<sup>1</sup>To whom correspondence should be addressed. E-mail: rockmore@cs.dartmouth.edu.



**Fig. 1.** Workflow of the analysis presented here. We begin by preprocessing our source images and training a set of basis functions to optimally represent them. We then convolve an authentic Bruegel and a randomly selected imitator image with the learned basis functions and estimate the kurtosis of the response distribution for each of these. We say that the image whose representation had higher mean kurtosis was more similar to the control set of authentic Bruegels than the other.

In the sparse coding model, basis functions are trained to maximize the kurtosis of the distribution of responses of the projection of random patches selected from images in the input space onto those basis functions. Such a model assumes that any grayscale image  $I(x, y)$  can be represented as a linear superposition of basis functions,  $\phi_i(x, y)$ , according to coefficients  $a_i$  (12):

$$I(x, y) = \sum_i a_i \phi_i(x, y).$$

Specifically, we derive for each set of images a corresponding set of functions that most sparsely represent those images, according to the following cost function:

$$E = -[\text{preserve information}] - \lambda[\text{sparseness of } a_i],$$

where  $\lambda$  is a positive constant that controls the importance of the sparseness constraint for the coefficients  $a_i$ . Information preservation is usually quantified as

$$-\sum_{x,y} \left[ I(x, y) - \sum_i a_i \phi_i(x, y) \right]^2.$$

Furthermore, the pairwise sparseness of the coefficients  $a_i$  is constrained by a function that heavily penalizes instances in which many coefficients simultaneously deviate from zero:

$$-\sum_i S\left(\frac{a_i}{\sigma}\right),$$

where  $\sigma$  is a scaling constant.  $S$  is a suitable nonlinear function, for example,  $\log(1 + x^2)$ , or the negative logarithm of a prior (sparse) distribution over the  $a_i$  (12).

## Methods

The Bruegel and Bruegel imitator source images were obtained from high-resolution scans of slides provided by the Metropolitan Museum of Art (MMA). In these images, a single pixel spans the same physical area.

Images were converted to (lossless) Portable Network Graphics format and converted to grayscale using the following transform:

$$I_{\text{gray}} = 0.2989 \times R + 0.5870 \times G + 0.1140 \times B,$$

where  $R$ ,  $G$ , and  $B$  are the red, green, and blue components of the image pixels, respectively, and  $I_{\text{gray}}$  is the resultant image. For each image, each dimension was then downsampled to one-half its original extent, yielding an image with four times fewer pixels. This was done to increase computational efficiency by reducing the size of the basis functions, whose size should roughly correspond to the size of salient features in the images. Throughout all stages of our preprocessing, the images were never compressed using a lossy compression algorithm.

Next, a random square section was taken from each image with side length equal to the smaller of the two image dimensions. This random square section became the source image for the images in each set used in our experiments. Random square sections from images in the Bruegel set have amplitude spectra that fall off as  $1/f$ . Because of this, the variance in certain directions in the input space is much larger than in others, which can create difficulties for the convergence of the sparse coding algorithm (19). Furthermore, because we extracted square patches from each image when training our basis functions, the sampling density in the high-frequency areas of the 2D frequency plane is higher than in other frequency bands (19).

To ameliorate these effects as well as those due to noise and aliasing artifacts inherent to downsampling, we used a special filter as described in ref. 19

$$R(f) = f e^{-(f/f_0)^n}$$

(where  $f_0 = \lfloor s/2 \rfloor$ , for  $s$ , the side length of the square image patch, and  $n = 4$ ) to “whiten” the images by flattening their amplitude spectra and attenuating high-frequency components. This also removes first-order correlations in the image, leaving only higher-order redundancies, the very statistical characteristics of the images that sparse coding is designed to exploit. Finally, the grayscale pixel values in each patch were linearly rescaled so that the maximum (absolute) pixel value was one.

In all experiments, three sets of basis functions, of sizes  $8 \times 8$ ,  $12 \times 12$ , and  $16 \times 16$  pixels, were randomly initialized to values on the interval  $[-0.5, 0.5]$ . In each set, the number of functions was set equal to the square of the side length of a particular function from that set (e.g., we trained  $64 \times 8$  bases). In a given training epoch, an image was selected at random and, from this image, 2048 random patches of the same size as the basis functions were selected.

A set of coefficients  $\{a_i\}$  was derived for each image patch according to the sparseness constraint described above. Using these coefficients, the set of basis functions was updated according to the following rule:

$$\Delta \phi_i(x_m, y_n) = \eta \langle a_i [I(x_m, y_n) - \hat{I}(x_m, y_n)] \rangle,$$

where  $\hat{I}(x_m, y_n)$  is the current image reconstruction (given  $\phi_i$  and  $a_i$ ) and  $\eta$  is the learning rate (in our experiments,  $\eta = 0.5$ ). Note that  $\langle \cdot \rangle$  denotes the averaging operator, in this case applied across all extracted image patches. Each set of basis functions was trained for 1,000 epochs.

In our experiments, we used the *sparsenet* software package, written and maintained by Bruno Olshausen (20). All model parameters were set to the default *sparsenet* parameters, except where indicated. Note also that the same image preprocessing steps were applied to the source authentic Bruegels used for model training, as well as to the authentic and imitator images used during testing.

## Results

To assess the robustness and validity of using the sparse coding model to quantify artistic style, we compared a set of established drawings by Pieter Bruegel the Elder with a set of well-known Bruegel imitations. There is consensus among art historians as to who actually created these works of art, providing us with a ground truth upon which to examine our approach to the quantification of artistic style (13). We use the measure of kurtosis to quantify similarity of style in the context of Bruegel’s work: If a set

of basis functions trained on a group of Bruegel drawings more sparsely represents one unknown than another, then we judge that unknown to be more like the works that were used to generate the functions in the first place. This follows directly from the assumptions of the model. Indeed, because the basis functions are trained exactly according to the constraint that sparseness should be maximized, we conclude that the sparsest representation indicates greatest similarity with respect to the learned features.

In our task, we compared an authentic Bruegel and a Bruegel imitator to the same set of authentic Bruegel images. Iteratively, we held out each authentic Bruegel and trained a set of basis functions on the remaining Bruegels. We then selected a random imitator image. We convolved 2,048 patches from each held-out image with each set of Bruegel basis functions, obtaining for each patch an estimate of the kurtosis of that distribution of responses. Once estimates were obtained, we compared the authentic drawing to the imitator's work in the following manner: We directly compared the means of the kurtosis distributions, asking if the mean of the estimated kurtosis distribution for the authentic Bruegel was greater than that of the imitator image's distribution. We performed this process 25 times for each authentic Bruegel, selecting a different random imitator image each time, yielding 625 pairwise comparisons per authentic Bruegel per basis set.

From this comparison-of-means test, we estimated the  $p$  value of observing the number of times the authentic Bruegel was judged more similar to the other authentic drawings than the random imitator, under the assumption that the two were drawn from the same distribution. This allows us to place concrete confidence estimates on the reliability of our predictions. In our analyses, we used basis functions of sizes  $8 \times 8$ ,  $12 \times 12$ , and  $16 \times 16$  pixels.

Table 1 shows the results of the comparison of the means of the kurtosis estimates for authentic and imitation Bruegel drawings across five runs. For seven of eight authentic Bruegels, the Bruegel work was considered more authentic than the imitator in more than 50% of trials at all three spatial scales. Furthermore, as Table 2 indicates, all of these results were statistically significant at the  $\alpha = 0.05$  level. One of the authentic drawings (MMA catalogue no. 11), however, was consistently judged less similar to the other secure Bruegels than a random imitator at two of three spatial scales. Nevertheless, the accuracy, stability, and robustness of our results indicate that kurtosis is an appropriate proxy for judging similarity in stylometric analysis. Indeed, we see by this measure that the excluded authentic Bruegel drawings were judged to be "more similar" to the other authentic Bruegels than the imitator drawings at a level that was consistently statistically significant.

We also asked whether each imitation could be successfully distinguished from the group of authentic Bruegels. To this end, we computed the  $p$  values of a nonparametric  $t$  test performed on the distributions of kurtosis values for *all* authentic Bruegels versus a distribution for each imitation. We estimated the kurtosis distribution for the authentic drawings by using a model trained on all of the authentic images, and then projected patches from each authentic image onto this basis and obtained

**Table 1. Classification of Bruegel drawings by comparison of means, specified by MMA catalogue number**

No.	$8 \times 8$	$12 \times 12$	$16 \times 16$
3	1.0	1.0	1.0
4	1.0	1.0	1.0
5	1.0	1.0	1.0
6	0.59	0.61	0.69
9	0.86	0.87	0.82
11	0.55	0.41	0.43
13	0.70	0.62	0.73
20	0.94	1.0	1.0

**Table 2.  $p$  values for the results in Table 1**

No.	$8 \times 8$	$12 \times 12$	$16 \times 16$
3	$p < 10^{-188}$	$p < 10^{-188}$	$p < 10^{-188}$
4	$p < 10^{-188}$	$p < 10^{-188}$	$p < 10^{-188}$
5	$p < 10^{-188}$	$p < 10^{-188}$	$p < 10^{-188}$
6	$p < 10^{-5}$	$p < 10^{-8}$	$p < 10^{-20}$
9	$p < 10^{-79}$	$p < 10^{-83}$	$p < 10^{-61}$
11	$p = 0.0125$	$p \approx 1.0$	$p \approx 0.9999$
13	$p < 10^{-24}$	$p < 10^{-8}$	$p < 10^{-31}$
20	$p < 10^{-129}$	$p < 10^{-188}$	$p < 10^{-188}$

the kurtosis distribution in the manner described above. For each imitation image, we obtained a kurtosis distribution by projecting patches of the image onto the same model trained on the authentic Bruegels. In all cases, distributions were estimated as the aggregate of distributions obtained over five independent runs. We performed this test, as before, for all three spatial scales. Four out of five imitations were judged significantly different from the authentic Bruegels at the  $\alpha = 0.05$  significance level at all three spatial scales. The remaining imitation was judged significantly different under the same criterion at two of three spatial scales. The first column of Table 3 gives the best  $p$  value obtained for each imitation at any spatial scale.

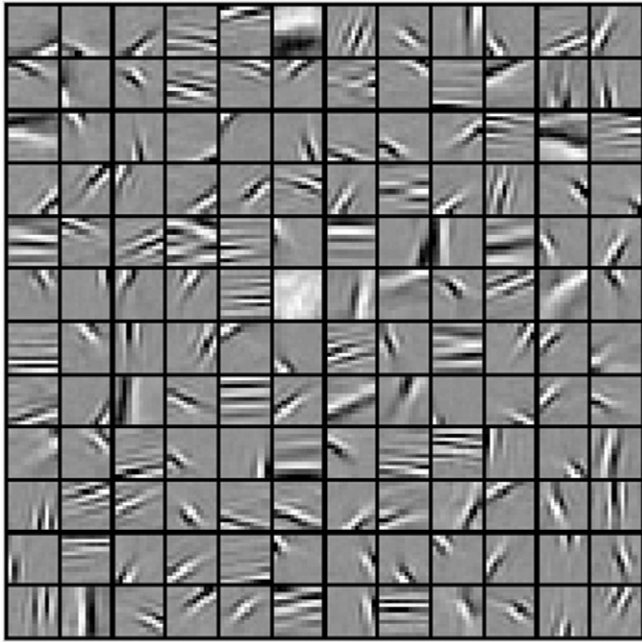
**Comparison with Quadrature Mirror Filters.** It is useful to compare our work with other multiscale approaches to the quantification of style. First, we consider the approach presented in ref. 2, which used the same set of authentic and imitation Bruegels and summarized the images with sets of feature vectors constructed from the marginal and error statistics (computed from the best linear predictor) derived from a multiscale analysis using quadrature mirror filters (QMFs) of a tiling of each image. This analysis maps each image to a "cloud" (set) of points in a 72-dimensional space whereupon the images are assigned a dissimilarity given by the (Hausdorff) distance between the point clouds.

Our analysis is similar in that we also project the images onto a set of functions to quantify their similarity, but differs from the method used in ref. 2 in at least two important ways. First, rather than use an arbitrarily chosen wavelet decomposition at fixed scales, we take advantage of the statistics of the image spaces themselves to train functions to optimally represent those spaces, according to the sparseness constraint. As shown in Fig. 2, our basis functions occur at several scales and are localized and band-pass in nature. Second, our notion of similarity is based on direct comparison of a single metric associated with a particular image, given a set of basis functions (i.e., kurtosis), rather than a set of quantities (like the statistics of a wavelet decomposition). In this sense, our approach can be considered simpler and more germane to the task.

The experiments in ref. 2 demonstrated the ability to distinguish the authentic and imitation images in three dimensions in an unquantified manner, by showing that multidimensional scaling of the distance matrix of authentic and imitation Bruegels produced a 3D representation in which the authentic Bruegel images were contained within a bounding sphere that did not contain any imitations. Our approach provides a direct method of

**Table 3. Best  $p$  values obtained for rejecting each imitation drawing, across all three approaches**

No.	Sparse coding	QMFs	Curvelets
1	$8.75 \times 10^{-8}$	1/9	2/9
2	$4.92 \times 10^{-198}$	1/9	1/3
3	$\approx 0$	1/9	1/3
4	$8.75 \times 10^{-57}$	1/9	2/9
5	$1.33 \times 10^{-99}$	1/9	2/9



**Fig. 2.** Set of 144  $12 \times 12$  basis functions trained on all authentic Bruegels for 1,000 epochs. Note that these functions are localized, oriented, and bandpass.

comparing a single statistic, making direct numerical comparison between our result and ref. 2 difficult. However, we can compare quantitatively the two approaches in an indirect manner by considering the probability of observing the imitation images in the context of each respective model.

Following (10) we can use a simple nonparametric statistic to rank sets of images, using only the summary statistics of the QMF image decompositions (see ref. 10 for further detail). According to this statistic, the lowest-ranking image in a set is most “unlike” the rest of the images in the group and thus obtains the lowest possible  $p$  value (see ref. 10 for further detail). We performed this test for each imitation with respect to the entire set of authentic Bruegel drawings, so that the smallest possible  $p$  value was one-ninth (eight authentic images and one imitation per test). In this test, each imitation was assigned the smallest possible  $p$  value (10) (see column 2 of Table 3).

**Comparison with Curvelets.** Curvelets (14, 15) provide another means for image analysis that appears a priori relevant for the analysis of visual art. Furthermore, they are also related to sparse coding techniques: Curvelets are effectively a formalization of the types of functions learned by sparse coding models like the one used in our experiments (14, 15). Motivated by this fact, we performed an experiment analogous to the one described in ref. 2, replacing the QMFs with curvelets, using the *CurveLab* software package (21). As in ref. 2, we took equally sized sections of the central  $512 \times 512$  pixel section of all authentic and imitation Bruegels for the sizes  $512 \times 512$ ,  $128 \times 128$ ,  $64 \times 64$ , and  $32 \times 32$  pixels (note that the *CurveLab* software did not support patch sizes below  $32 \times 32$  pixels). We represented each section using a vector of coefficient statistics (the first four moments of the distribution of coefficients) at each subband (across scale and orientation). At each of the four spatial scales, we computed the Hausdorff distance between the sets of statistics for all pairs of images. We then performed both metric and nonmetric multi-dimensional scaling on the resultant distance matrix.

Unlike ref. 2, at none of these spatial scales was there an implicit grouping of the images that allowed us to separate

the authentic drawings from the imitations in three or fewer dimensions. However, we were able to achieve classification performance on par with our primary result (Table 1) by using a nearest neighbor approach (22) to classify points as authentic or imitation drawings, correctly identifying seven of eight authentic Bruegels as most similar to other authentic Bruegels, for the Hausdorff distance matrix obtained using  $32 \times 32$  pixel sections. Although this classification performance is equal to our own, we cannot estimate a  $p$  value for this process or directly evaluate its significance.

However, we performed the test described in ref. 10, as above, using the distance matrix between images obtained from the summary statistics of curvelet representations of images. In none of our tests, across any of the four spatial scales considered, was any imitation judged *least* like the authentic drawings (i.e., no imitation was ever assigned the lowest possible  $p$  value). Table 3 compares these numbers with those obtained for the two previous approaches detailed here, using  $p$  values obtained for the  $32 \times 32$  spatial scale. Note that, in making this comparison, we acknowledge that the test described in ref. 10 does not have the ability to give stronger measures of statistical significance. Nevertheless, our approach provides robust and strongly significant results.

From these experiments, we draw two important conclusions. First, sparse coding provides a method for classifying authentic Bruegel drawings (i.e., including authentic Bruegels with other authentic images) that is both robust and produces statistically significant results. Although curvelets produced a promising result on par with ours in the same task, the classification approach used here does not have a significance criterion associated with it. Finally, in the task of discriminating imitations from authentic Bruegels (i.e., rejecting an imitation as being sufficiently unlike the authentic images), we have shown that our method produces a result (a set of  $p$  values) that is numerically superior to the other two approaches presented.

## Discussion

We have shown that our model generates robust conclusions in line with art historical evidence in the analysis of the Bruegel drawings presented here. However, a few caveats bear mentioning. First, sparse coding does not necessarily lend itself to easy application in every analysis of artistic style. The current scenario dealt with specific examples of images in which the ratio of unknown to known exemplars was relatively small (per experimental run). Indeed, the validity of the sparse coding model, which by design attempts to generate a statistically representative set of basis functions for a given image space, rests on possessing a sufficient number of examples of that image space.

Also, as demonstrated by our analysis, the scale at which sparse coding analysis is performed can have a significant effect on the outcome. This results from several factors, including the actual physical area of each pixel, the size of salient features represented by the model, and the extent to which the images are down-sampled, among others.

Finally, any classification problem should be well posed and fairly constrained. For example, a sparse coding approach would be inappropriate when attempting to compare works from artists whose styles vary considerably. The technique is most apt to deal with situations where the notion of stylistic similarity can be reduced to straightforward assertions, e.g., comparing a known work by a particular artist to sets of works by the same artist, or attempting to authenticate a work by comparing it to known works by a particular artist. Additionally, some care should be taken in choosing works that contain similar subject matter, such as landscapes, if this is important to the question at hand.

## Conclusion

We have presented the use of sparse coding as a unique technique for the quantification of artistic style. We applied it to the significant art historical problem of authentication and, in particular, to classify authentic Bruegel drawings, as well as to distinguish imitation Bruegel drawings from secure ones. Our classification technique relied on a straightforward comparison of a single statistic, making it an attractive alternative to wavelet-like techniques that may involve a more complicated statistical framework. Furthermore, we have shown that our technique produces a numerically superior result to two alternative approaches at separating imitations from the set of authentic Bruegel drawings. The superior performance of the sparse coding model may result from the fact that it creates an adaptive representation of the image space, an attribute that more restrictive methods (e.g., QMFs) do not possess. This example is both evidence for the

utility of sparse coding for authentication as well as a validation of the general concept.

Although this success is encouraging, we see sparse coding and other digital stylometric tools as useful supplements to, rather than replacements for, traditional tools and techniques in art historical analysis. These digital techniques can assist art historians in making judgements and may provide detailed information about subtleties inherent to a particular artist's style that are not immediately observable.

**ACKNOWLEDGMENTS.** Thanks to E. Postma, J. Coddington, R. Johnson, M. Albert, B. Olshausen, D. Field, G. Leibon, and P. Kostelec for helpful conversations. We gratefully acknowledge the support of National Science Foundation Grant DMS-0746667 and a grant from the William H. Neukom 1964 Institute for Computational Science.

1. Holmes DJ, Kardos J (2003) Who was the author? An introduction to stylometry. *Chance*, 16(2):5–8.
2. Lyu S, Rockmore DN, Farid H (2004) A digital technique for art authentication. *Proc Natl Acad Sci USA*, 101(49):17006–17010.
3. Taylor R, Micolich A, Jonas D (1999) Fractal analysis of Pollock's drip paintings. *Nature*, 399:422.
4. Berezhnoy I, Postma E, van den Herik J (2007) Computer analysis of van Gogh's complementary colours. *Pattern Recogn Lett*, 28:703–709.
5. Taylor R, et al. (2007) Authenticating Pollock paintings using fractal geometry. *Pattern Recogn Lett*, 28:695–702.
6. Graham DJ, Friedenbergr JD, Rockmore DN, Field DJ (2009) Mapping the similarity space of paintings: Image statistics and visual perception. *Visual Cognition* doi: 10.1080/13506280902934454.
7. Redies C (2007) A universal model of esthetic perception based on the sensory coding of natural stimuli. *Spatial Vision*, 21:97–117.
8. Graham DJ, Friedenbergr J, Rockmore DN (2009) Efficient visual system processing of spatial and luminance statistics in representational and non-representational art. *PSoc Photo-Opt Ins*, 7240:1N1–1N10.
9. Mureika J (2005) Fractal dimensions in perceptual color space: a comparison study using Jackson Pollock's art. *Chaos*, 15(4):043702 doi: 10.1063/1.2121947.
10. Rockmore D, Leibon G (2007) *Matematica e Cultura 2007. In cammino verso l'autenticazione digitale* (Springer-Verlag, Milano, Italy).
11. Johnson CR, Jr, et al. (2008) Image processing for artist identification—computerized analysis of Vincent van Gogh's painting brushstrokes. *IEEE Signal Proc Mag, Special Issue on Visual Cultural Heritage*, 25(4):37–48.
12. Olshausen B, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609.
13. (2001) *Pieter Bruegel—Drawings and Prints*, ed NM Orenstein (Metropolitan Museum of Art, New York, Yale Univ Press, New Haven, CT).
14. Candès EJ, Donoho DL (1999) Curvelets—a surprisingly effective nonadaptive representation for objects with edges. *Curves and Surfaces*, ed LL Schumaker (Vanderbilt Univ Press, Nashville, TN).
15. Donoho DL, Flesia AG (2001) Can recent innovations in harmonic analysis 'explain' key findings in natural image statistics?. *Network-Comp Neural*, 12:371–393.
16. Hyvärinen A, Hoyer PO, Hurri J (2003) Extensions of ICA as models of natural images and visual processing. *Proceedings of the Fourth International Symposium on Independent Components Analysis and Blind Signal Separation* pp 963–974.
17. Olshausen BA, Field DJ (2004) Sparse coding of sensory inputs. *Curr Opin Neurobiol*, 14(4):481–487.
18. Graham DJ, Field DJ (2007) Statistical regularities of art images and natural scenes: Spectra, sparseness and nonlinearities. *Spatial Vision*, 21:149–164.
19. Olshausen B, Field DJ (1997) Sparse coding with an overcomplete basis set: A strategy employed by v1?. *Vision Res*, 37(23):3311–3325.
20. Olshausen B *Sparsenet software package* Available at <https://redwood.berkeley.edu/bruno/sparsenet/>.
21. Demanet L *CurveLab software package* Available at <http://www.curvelet.org>.
22. (2005) *Nearest-Neighbor Methods in Learning and Vision*, eds G. Shakhnarovich, P. Indyk, and T. Darrell (MIT Press, Cambridge, MA).