Informational Aesthetics Measures

Jaume Rigau, Miquel Feixas, and Mateu Sbert ■ *University of Girona, Spain*

The Birkhoff aesthetic measure of an object is the ratio between order and complexity. Informational aesthetics describes the interpretation of this measure from an information-theoretic perspective. From these ideas, the authors define a set of ratios based on information theory and Kolmogorov complexity that can help to quantify the aesthetic experience.

In 1928, George D. Birkhoff formalized the aesthetic measure of an object as the quotient between order and complexity (see also the "Related Work" sidebar).[1] From Birkhoff's work, Max Bense,[2] together with Abraham Moles,[3] developed *informational aesthetics* (or *information-theoretic aesthetics* from the original German term), which defines the concepts of order and complexity from Shannon's notion of information.[4] As Birkhoff stated, formalizing these concepts, which depend on the context, author, observer, and so on, is difficult. Scha and Bod claimed that in spite of these measures' simplicity, "if we integrate them with other ideas from perceptual psychology and computational linguistics, they may in fact constitute a starting point for the development of more adequate formal models."[5]

The creative process generally produces order from disorder. Bense proposed a general schema that characterizes artistic production by the transition from the repertoire to the final product. He assigned a complexity to the repertoire, or palette, and an order to the distribution of its elements on the artistic product.

This article, an extended and revised version of earlier work,[6] presents a set of measures that conceptualizes Birkhoff's aesthetic measure from an informational viewpoint. These measures describe complementary aspects of the aesthetic experience and are normalized for comparison. We show the measures' behavior using three sets of paintings representing different styles that cover a representative feature range: from randomness to order. Our experiments show that both global and compositional measures extend Birkhoff's measure and help us understand and quantify the creative process.

Information theory and Kolmogorov complexity

Some basic notions of information theory,[4] Kolmogorov complexity,[7] and physical entropy[8] serve as background for our work.

Information-theoretic measures

Information theory deals with information transmission, storage, and processing.[4] Researchers in fields such as physics, computer science, statistics, biology, image processing, and learning use information theory.

Let $\mathcal{X}$ be a finite set and X be a random variable taking values x in $\mathcal{X}$ with distribution $p(x) = Pr[X = x]$ (that is, the probability that variable X takes value x). Likewise, let Y be a random variable taking values y in $\mathcal{Y}$. We characterize an information channel $X \rightarrow Y$ between two random variables (input X and output Y) by a probability transition matrix that determines the output distribution given the input.

We define the *Shannon entropy* $H(X)$ of a random variable X by

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

The Shannon entropy $H(X)$, also denoted by $H(p)$, measures the average uncertainty of random variable X and fulfills $0 \le H(X) \le \log |\mathcal{X}|$. If the logarithms are taken in base 2, we express entropy in bits.

The *conditional entropy* is defined by

$$H(X|Y) = -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x,y) \log p(x|y)$$

Published by the IEEE Computer Society

Related Work in Informational Aesthetics

Eighty years ago, Birkhoff formalized the notion of beauty by introducing the *aesthetic measure*, defined as the ratio between order and complexity.[1] According to this measure, "the complexity is roughly the number of elements that the image consists of and the order is a measure for the number of regularities found in the image."[2]

Birkhoff suggested that aesthetic feelings stem from the harmonious interrelations inside the object and that the aesthetic measure is determined by the order relations in the object. He identified three successive phases in the aesthetic experience:

- A preliminary effort of attention, which is necessary for the act of perception and increases proportionally to the object's complexity (*C*).
- The feeling of value or aesthetic measure (*M*) coming from this effort.
- The verification that the object is characterized by certain harmony, symmetry, or order (*O*), which seems necessary for the aesthetic effect.

From these considerations, Birkhoff defined the aesthetic measure as $M = O/C$.

Birkhoff understood the impossibility of comparing objects of different classes and accepted that the aesthetic experience depends on the observer. So, he proposed restricting the group of observers and applying the measure only to similar objects.

Using information theory, Bense proposed both the redundancy and Shannon entropy to quantify, respectively, an artistic object's order and complexity.[3] According to Bense, any artistic creation process involves a determined repertoire of elements (such as colors, sounds, and phonemes) that is transmitted to the final product. The creative process is selective (that is, to create is to select). For instance, if the repertoire is given by a palette of colors with a probability distribution, the final product (in our case, a painting) is a selection (a realization) of this palette on a canvas. Although the distribution of elements of an aesthetic state has a certain order, the repertoire shows a certain complexity. Bense also distinguished between a global complexity, formed by partial complexities, and a global order, formed by partial orders.

Other authors have also introduced measures to quantify aesthetics. Koshelev considered that the running time $t(p)$ of a program p that generates a given design is a formalization of Birkhoff's complexity *C*. In addition, a monotonically decreasing function of the program's length $l(p)$ (that is, Kolmogorov complexity) represents Birkhoff's order *O*.[4] So, looking for the most attractive design, $M = 2^{-l(p)}/t(p)$ defines the aesthetic measure. Machado and Cardoso established that an aesthetic visual measure depends on the ratio between image complexity and processing complexity.[5] They estimated both using real-world compressors (JPEG and fractal, respectively). They considered that images that are simultaneously visually complex and easy to process have a higher aesthetic value.

Greenfield[6] and Hoenig[7] provide excellent overviews of the history of the aesthetic measures.

References

1. G.D. Birkhoff, *Aesthetic Measure*, Harvard Univ. Press, 1933.
2. R. Scha and R. Bod, "Computationele Esthetica," (Computational Esthetics), *Informatie en Informatiebeleid*, vol. 11, no. 1, 1993, pp. 54-63; English translation available at http://iaaa.nl/rs/compestE.html.
3. M. Bense, *Einführung in die informationstheoretische Asthetik. Grundlegung und Anwendung in der Texttheorie* (Introduction to the Information-theoretical Aesthetics. Foundation and Application to the Text Theory), Rowohlt Taschenbuch Verlag, 1969.
4. M. Koshelev, "Towards the Use of Aesthetics in Decision Making: Kolmogorov Complexity Formalizes Birkhoff's Idea," *Bull. European Assoc. Theoretical Computer Science*, vol. 66, Oct. 1998, pp. 166-170.
5. P. Machado and A. Cardoso, "Computing Aesthetics," *Advances in Artificial Intelligence, Proc. 14th Brazilian Symp. Artificial Intelligence* (SBIA 98), LNCS 1515, Springer, 1998, pp. 219-228.
6. G. Greenfield, "On the Origins of the Term 'Computational Aesthetics,'" *Proc. Eurographics Workshop Computational Aesthetics in Graphics, Visualization, and Imaging*, Eurographics Assoc., 2005, pp. 9-12.
7. F. Hoenig, "Defining Computational Aesthetics," *Proc. Eurographics Workshop Computational Aesthetics in Graphics, Visualization, and Imaging*, Eurographics Assoc., 2005, pp. 13-18.

where $p(x, y) = Pr[X = x, Y = y]$ is the joint probability, and $p(x|y) = Pr[X = x|Y = y]$ is the conditional probability. The conditional entropy $H(X|Y)$ measures the average uncertainty associated with X if we know the outcome of Y. The *mutual information* between X and Y is defined by $I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$, and represents the shared information between X and Y.

The Shannon source-coding theorem is a fundamental result of information theory. This theorem encodes an object to store or transmit it efficiently. The theorem expresses that an optimal code's minimal length (for instance, a Huffman code) fulfills

$$H(X) \leq \overline{\ell} < H(X) + 1 \qquad (1)$$

where $\overline{\ell}$ is the expected length of the optimal binary code for X.

Another interesting property of the entropy is the *Jensen-Shannon inequality*, which is expressed by

$$JS(\pi_1,\ldots,\pi_n;p_1,\ldots,p_n) \equiv H(\sum_{i=1}^{n}\pi_i p_i) - \sum_{i=1}^{n}\pi_i H(p_i) \geq 0 \quad (2)$$

where $JS(\pi_1, \ldots, \pi_n; p_1, \ldots, p_n)$ is the *Jensen-Shannon divergence* of probability distributions $p_1, \ldots, p_n$ with n prior probabilities or weights $\pi_1, \ldots, \pi_n$; fulfilling $\Sigma_{i=1}^{n}\pi_i = 1$. The Jensen-Shannon divergence measures how far the probabilities p_i are from their likely joint source $\Sigma_{i=1}^{n}\pi_i p_i$ and equals zero if, and only if, all p_i are equal.

Kolmogorov complexity and the similarity metric

The *Kolmogorov complexity* $K(x)$ of a string x is the length of the shortest program to compute x on an appropriate universal computer.[6] Essentially, a string's Kolmogorov complexity is the length of its ultimate compressed version and is machine-independent up to an additive constant. The conditional complexity $K(x|y)$ of x relative to y is defined as the length of the shortest program to compute x given y as an auxiliary input to the computation. The joint complexity $K(x, y)$ represents the length of the shortest program for the pair (x, y). The Kolmogorov complexity is also called algorithmic information or algorithmic randomness.

Information distance is defined as the length of the shortest program that computes x from y and y from x.[7] Up to an additive logarithmic term, the information distance is given by $E(x, y) = \max\{K(y|x), K(x|y)\}$. This measure is a metric. Long strings that differ by a small amount are intuitively closer than short strings that differ by the same amount. Hence, the necessity to normalize the information distance arises. Li and colleagues[7] define a normalized version of $E(x, y)$, called the *normalized information distance* or the *similarity metric*:

$$NID(x,y) = \frac{\max\{K(x|y),K(y|x)\}}{\max\{K(x),K(y)\}} = \frac{K(x,y) - \min\{K(x),K(y)\}}{\max\{K(x),K(y)\}} \quad (3)$$

NID is also a metric and takes values in [0, 1]. It's universal in the sense that if two strings are similar according to the feature described by a particular normalized admissible distance (not necessarily a metric), they're also similar in the sense of the normalized information metric. Because of the Kolmogorov complexity's noncomputability, a feasible version of *NID*, called *normalized compression distance*, is defined as

$$NCD(x,y) = \frac{C(x,y) - \min\{C(x),C(y)\}}{\max\{C(x),C(y)\}} \quad (4)$$

where $C(x)$ and $C(y)$ represent the length of compressed string x and y, respectively, and $C(x, y)$ the length of the compressed pair (x, y). Therefore, *NCD* approximates *NID* by using a standard real-world compressor.

Physical entropy

Looking at a system from an observer's angle, Zurek[8] defined the *physical entropy* as the sum of the missing information (Shannon entropy) and the algorithmic information content (Kolmogorov complexity) of the available data:

$$S_d = H(X_d) + K(d) \quad (5)$$

where d is the system's observed data, $K(d)$ is the Kolmogorov complexity of d, and $H(X_d)$ is the conditional Shannon entropy or our ignorance about the system given d.

Physical entropy reflects the fact that measurements increase our knowledge about a system. In the beginning, we have no knowledge about the system's state, so the physical entropy reduces to the Shannon entropy, reflecting our total ignorance. If the system is in a regular state, physical entropy decreases as we make more measurements. In this case, we increase our knowledge about the system and might be able to efficiently compress the data. If the state isn't regular, we can't achieve compression, and the physical entropy remains high. According to Zurek, we can view this compression process from the perspective of an information-gathering and using system entity, such as a Maxwell's demon, capable of measuring and modifying its strategies based on the measurements' outcomes.

Global aesthetic measures

We consider three basic concepts of Bense's creative process:

- the *initial repertoire*—the basic states (in our case, a wide range of colors that we assume are finite and discrete);
- the *used palette* (selected repertoire)—the range of colors selected by the artist with a given probability distribution; and
- the *final color distribution*—the arrangement of the palette colors on a physical support (canvas).

Our set of measures uses these concepts to extend Birkhoff's measure using information theory and Kolmogorov complexity.

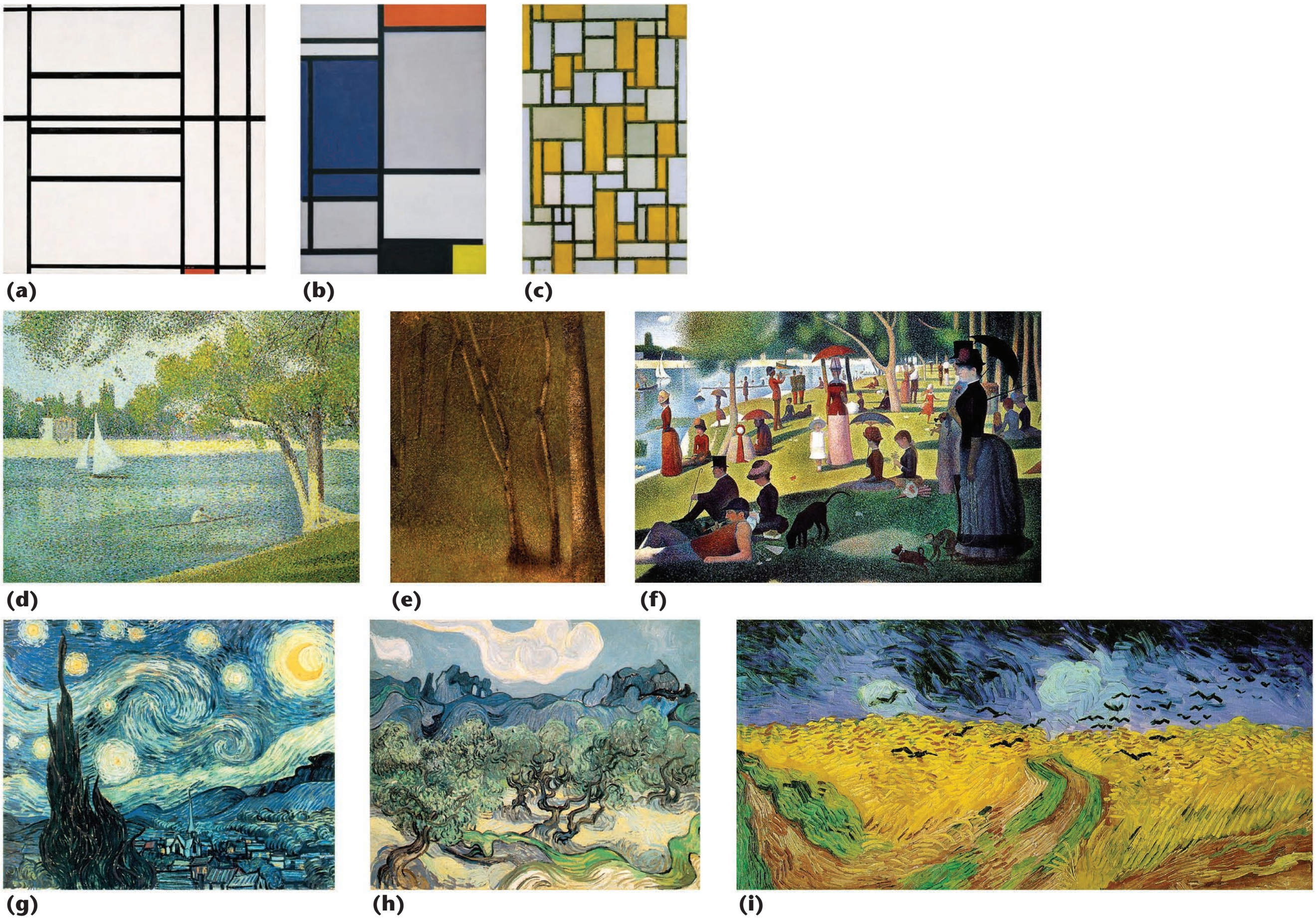


Figure 1. Paintings used in our tests. (a) Composition with Red, Piet Mondrian, 1938–1939; (b) Composition with Red, Blue, Black, Yellow, and Gray, Piet Mondrian, 1921; (c) Composition with Grid 1, Piet Mondrian, 1918; (d) The Seine at Le Grande Jatte, Georges-Pierre Seurat, 1888; (e) Forest at Pontaubert, Georges-Pierre Seurat, 1881; (f) Sunday Afternoon on the Island of La Grande Jatte, Georges-Pierre Seurat, 1884-1886; (g) The Starry Night, Vincent van Gogh, 1889; (h) Olive Trees with the Alpilles in the Background, Vincent van Gogh, 1889; and (i) Wheat Field under Threatening Skies, Vincent van Gogh, 1890.

For a given color image I of N pixels, we use an sRGB color representation based on a repertoire of 256^3 colors ($\mathcal{X}_{rgb}$). We reduce the $\mathcal{X}_{rgb}$ range using the luminance Y_{709} ($\mathcal{X}_\ell = [0, 255]$). From the normalization of the intensity histograms of $\mathcal{X}_{rgb}$ and $\mathcal{X}_\ell$, using $256^3(N_b^{rgb})$ and $256(N_b^\ell)$ bins, respectively, we obtain the probability distributions of the random variables X_{rgb} and X_ℓ. The maximum entropy H_{max} for these random variables is $\log|N_b^{rgb}| = 24$ and $\log|N_b^\ell| = 8$, respectively.

Throughout this article, we use the following notions:

- a palette (X_{rgb} or X_ℓ), given by the image's normalized intensity histogram;
- the palette entropy or pixel uncertainty (H_p), obtained from $H(X_{rgb})$ or $H(X_\ell)$;
- the image information content or image uncertainty (NH_p); and
- an image's Kolmogorov complexity (K).

We applied our measures to the set of paintings shown in Figure 1. Table 1 (next page) lists their sizes as well as the size and compression ratio achieved by the JPEG compressor.

Shannon's perspective

Bense proposed using redundancy to measure order in an aesthetic object (see the "Related Work" sidebar on page 25). When we apply this idea to an image or painting, the absolute redundancy $H_{max} - H_p$ expresses the reduction of uncertainty due to the choice of a palette with a given color probability distribution instead of a uniform distribution. Thus, we can express the aesthetic measure as the relative redundancy:

$$M_B = \frac{H_{max} - H_p}{H_{max}}$$

From a coding perspective, this measure represents the gain from using an optimal code to compress the image (Equation 1). The redundancy expresses one aspect of the creative process: the artist's selected

Table 1. Size of the original files and size and compression ratio for the paintings in Figure 1, using JPEG compression with the maximum quality option.

	Original image file		Compressed file	
Painting	**Pixels**	**Bytes**	**Bytes**	**Ratio**
Mondrian-1 (a)	316,888	951,862	160,557	5.928
Mondrian-2 (b)	139,050	417,654	41,539	10.055
Mondrian-3 (c)	817,740	2,453,274	855,074	2.869
Seurat-1 (d)	844,778	2,535,422	1,473,336	1.721
Seurat-2 (e)	857,540	2,572,674	1,530,889	1.681
Seurat-3 (f)	375,750	1,128,306	519,783	2.171
Van Gogh-1 (g)	831,416	2,495,126	919,913	2.712
Van Gogh-2 (h)	836,991	2,511,850	862,274	2.913
Van Gogh-3 (i)	856,449	2,570,034	1,203,527	2.135

Table 2. Entropy $H(X_{rgb})$ and global aesthetic measures M_B, M_K, and M_Z for the paintings in Figure 1.

		Aesthetic measures		
Painting	$H(X_{rgb})$	M_B	M_K	M_Z
Mondrian-1 (a)	8.168	0.660	0.831	0.504
Mondrian-2 (b)	9.856	0.589	0.900	0.758
Mondrian-3 (c)	14.384	0.401	0.651	0.418
Seurat-1 (d)	14.976	0.376	0.419	0.068
Seurat-2 (e)	18.180	0.243	0.405	0.214
Seurat-3 (f)	17.045	0.290	0.539	0.351
van Gogh-1 (g)	17.204	0.283	0.631	0.485
van Gogh-2 (h)	17.288	0.280	0.657	0.523
van Gogh-3 (i)	17.689	0.263	0.532	0.364

palette. Table 2 shows significant differences in the M_B values for the set of paintings in Figure 1. To obtain these results, we computed a pixel's entropy using $H_p = H(X_{rgb})$ (thus, $H_{max} = 24$). From Mondrian-1 (Figure 1a) to van Gogh-3 (Figure 1i), the results reflect the high color homogeneity in Mondrian's paintings and the major color diversity in Seurat's and van Gogh's paintings. This measure only reflects the palette information and doesn't account for colors' spatial distribution on canvas. Thus, the geometry (Mondrian), pointillism's randomness (Seurat), and landscape elements (van Gogh and Seurat) are compositional features perceived by a human observer but not captured by M_B. The measures described in the following sections address these features.

Kolmogorov's perspective

From a Kolmogorov complexity perspective, we can measure the order in an image by the difference between the image size (obtained using a constant length code for each color) and its Kolmogorov complexity. This corresponds to the space saving defined as the size reduction relative to the uncompressed size. The order's normalization gives us the aesthetic measure:

$$M_K = \frac{NH_{max} - K}{NH_{max}}$$

M_K takes values in [0, 1] and expresses the image's degree of order without any prior knowledge of the palette (the higher the image's degree of order, the higher the compression ratio). Because of K's noncomputability, we use real-world compressors to estimate it (that is, we approximate K's value by the size of the corresponding compressed file). A compressor exploits both the selected palette's degree of order and the color position in the canvas. We selected the JPEG compressor because of its ability to discover patterns, in spite of (or thanks to) losing information that's imperceptible by the human eye. This is closer to the aesthetic experience than using lossless compressors, which usually have lower compression ratios so keep all the original information, including information that human observers can't distinguish. Nevertheless, to avoid losing significant information, we

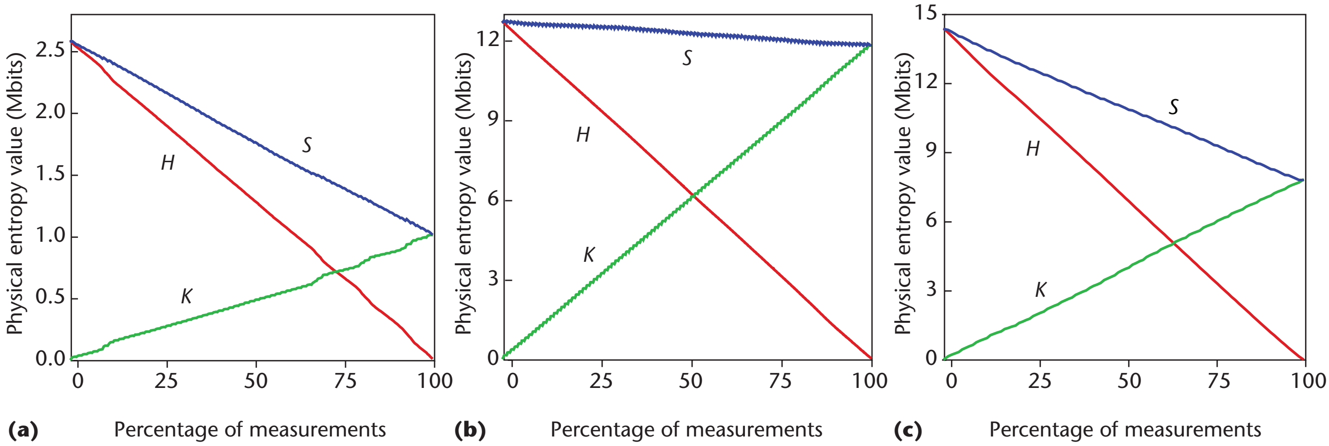


Figure 2. The evolution of physical entropy (*S*) (missing information *H* + Kolmogorov complexity *K*) for three paintings shown in Figure 1. The missing information is captured by $H_p = H(X_{rgb})$ and the Kolmogorov complexity has been approximated using the JPEG compressor. (a) Mondrian-1 (Figure 1a), (b) Seurat-1 (Figure 1d), and (c) van Gogh-1 (Figure 1g).

use a JPEG compressor with the maximum quality option (see Table 1).

For the results in Table 2, we calculated M_K using $H_{max} = 24$. Although a strict ordering on M_K values mixes paintings of different artists, the averages of the three sets of paintings are clearly separate. In descending order, the groups are Mondrian, van Gogh, and Seurat. The pairs of paintings (Mondrian-3, van Gogh-2) and (van Gogh-3, Seurat-3) have similar M_K values. This is probably because the compressor can detect more homogeneity (or heterogeneity) than the human eye. For instance, the interior of some regions in the Mondrian-3 painting is more heterogeneous than it appears at first glance.

Frieder Nake, a Bense disciple and pioneer in algorithmic art (that is, art explicitly generated by an algorithm), considered a painting as a hierarchy of signs, where at each level of the hierarchy we could determine the statistical information content. He conceived the computer as a universal picture generator capable of "creating every possible picture out of a combination of available picture elements and colors."[9] Nake's theory of algorithmic art fits well with Kolmogorov's perspective, because you can consider a painting's Kolmogorov complexity as the length of the shortest program generating it.

Zurek's perspective

We developed a new version of Birkhoff's measure based on Zurek's physical entropy.[8] Zurek's work lets us look at the creative process as an evolutionary process from the initial uncertainty (Shannon entropy) to the final order (Kolmogorov complexity). We can interpret this approach as a transformation of the color palette's initial probability distribution to the algorithm describing the final painting.

Inspired by physical entropy (Equation 5), we define a measure given by the ratio between the reduction of uncertainty (because of the compression achieved by Kolmogorov complexity) and the image's initial information content. Assuming that each pixel's Shannon entropy times the number of pixels (NH_p) gives an image's information content, we have

$$M_Z = \frac{NH_p - K}{NH_p}$$

This normalized ratio quantifies the degree of order created from a given palette.

For Table 2, we computed M_Z using the JPEG compressor, $H_p = H(X_{rgb})$, and $H_{max} = 24$. Taking the average of M_Z for each artist gives us the same ordering as in the previous measure M_K. The low values for Seurat's paintings are due to their low compression ratio because of the pointillist style (see Table 1).

The plots in Figure 2 express, for three paintings, the physical entropy's evolution as we take more measurements. To simulate this evolution, we progressively discover each painting's content (columns from left to right), reducing the missing information (Shannon entropy) and compressing the discovered information (Kolmogorov complexity). The Mondrian paintings show on average a greater order than the van Gogh paintings, and the van Gogh paintings more than the Seurat paintings. So, we can more efficiently compress or comprehend our progressive knowledge about the paintings in the Mondrian case than in the other cases.

Quantifying the creative process. We can understand the global measures from the initial repertoire's

Table 3. The compositional aesthetic measures M_j, M_k, and M_s for the set of paintings in Figure 1 computed for $n = 16$.

Painting	$H(X_\ell)$	Aesthetic measures M_j	M_K	M_s
Mondrian-1 (a)	5.069	0.900	0.312	0.166
Mondrian-2 (b)	6.461	0.762	0.335	0.352
Mondrian-3 (c)	7.328	0.969	0.198	0.060
Seurat-1 (d)	7.176	0.984	0.161	0.025
Seurat-2 (e)	7.706	0.979	0.147	0.032
Seurat-3 (f)	7.899	0.960	0.164	0.055
van Gogh-1 (g)	7.858	0.953	0.179	0.070
van Gogh-2 (h)	7.787	0.948	0.170	0.074
van Gogh-3 (i)	7.634	0.957	0.159	0.057

complexity (logarithm of the number of repertoire states), the selected palette (Shannon entropy), and the final distribution (Kolmogorov complexity). From these complexities, we obtain the order, measuring the differences between them:

- in M_B, $H_{\max} - H_p$ is the palette redundancy;
- in M_K, $NH_{\max} - K$ is the compression achieved from the product's order; and
- in M_Z, $NH_p - K$ is the reduction of uncertainty produced while observing or recognizing the final product.

These differences quantify the creative process: the first represents the selection process from the initial repertoire, the second captures the order in the color distribution, and the third expresses the transition from the palette to the artistic object.

Compositional aesthetic measures

Bense considered the creative act a transition process from an initial repertoire to the distribution of its elements on the physical support (such as a canvas). Here, we introduce measures to analyze an image's composition (that is, the spatial distribution of colors from a given palette).

Order as self-similarity

To analyze an image's composition, the measures used must quantify the degree of correlation or similarity between image parts. The Jensen-Shannon divergence and the similarity metric can capture the spatial order.

Shannon's perspective. From Shannon's viewpoint, we can compute the similarity between an image's parts using the Jensen-Shannon divergence (Equation 2), which is a measure of discrimination between probability distributions. We can use this divergence to calculate the dissimilarity between diverse regions' intensity histograms. Thus, for a given decomposition of an image, the Jensen-Shannon divergence will quantify the spatial heterogeneity.

Although the ratio between the image's Jensen-Shannon divergence and the initial uncertainty H_p expresses the degree of dissimilarity, we define its complementary value as a measure of self-similarity:

$$M_j(n) = 1 - \frac{JS(\pi_1,\ldots,\pi_n; p_1,\ldots,p_n)}{H_p} = \frac{\sum_{i=1}^{m} \pi_i H(p_i)}{H_p}$$

where n is the resolution level (that is, number of regions desired), π_i is the area of region i, p_i represents the probability distribution of region i, and $H(p_i)$ is its entropy. The self-similarity measure takes values in [0, 1], decreasing the value with a finer partition. For a random image and a coarse resolution, the value should be close to 1.

Table 3 shows the values of M_j for the set of paintings. In our tests, we decomposed the paintings in a 4×4 regular grid and computed the histograms using the luminance Y_{709}. The high similarity between the palettes of the parts of a Seurat painting fits with the high values of M_j. On the other hand, Mondrian-2's lower self-similarity is due to the presence of regions with different palettes.

Kolmogorov's perspective. To measure the similarity between two parts of an image, we use the normalized information distance (Equation 3). As we described earlier, the information distance between two subimages is the length of the shortest program needed to transform the two subimages into each other. If we consider an image's degree of order as the self-similarity, we can measure it from the average *NID* between each subimage pair:

$M_k(n) = 1 - \text{avg}_{1 \le i < j \le n}\{NID(i, j)\}$

where n is the number of regions or subimages provided by a given decomposition, and $NID(i, j)$ is the distance between subimages I_i and I_j. This value ranges from 0 to 1 and expresses the degree of order inside the image.

For Table 3, we calculate the values of M_k for the set of paintings using a 4 × 4 regular grid and $NCD(i, j)$ (Equation 4) as an approximation of $NID(i, j)$. For our case, we computed the values of $C(I_i)$ and $C(I_i, I_j)$ in NCD ignoring the rest of the canvas information (that is, zero luminance in $I - I_i$ and $I - I_i - I_j$, respectively). As in the previous compositional measure, M_j, we classified the paintings according to the artist, but in reverse order. This is because, whereas M_j only measures the similarity between regions' palettes, M_k also measures the spatial distribution similarity of the palettes on the canvas.

Interpreting Bense's channel

We can further understand the creative process described by Bense as the realization of an information channel between the palette and the image's regions.

From a Shannon perspective, we present an algorithm that progressively partitions the image, extracting all its information until the painting is completely revealed. The information extraction's rate will depend on the painting's degree of order. For instance, if a painting was created by randomly distributing the colors on the canvas, any possible partition will obtain a small information gain. However, if the painting shows a certain degree of structure, we'll probably find a partition that will give us a larger information gain.

We construct this partitioning algorithm from an information channel $B \rightarrow R$ between the random variables B (input) and R (output), which represent, respectively, the set of intensity bins B and the set of regions R of an image. A conditional probability matrix defines this channel. This matrix expresses how the pixels corresponding to each intensity bin are distributed in the image's regions. Given N_b intensity bins and N_r regions in the image I of N pixels, the channel's three basic elements comprise

- The conditional probability matrix $p(R|B)$, which represents the transition probabilities from each bin of the histogram to the image's regions, is defined by $p(r_j|b_i) = n_{ij}/n_i$, where n_{ij} is the number of pixels of b_i into the region r_j, and n_i is the number of pixels of b_i. Conditional probabilities fulfill

$$\forall b \in \mathrm{B}. \sum_{j=1}^{N_r} p(r_j|b) = 1$$

- The input distribution $p(B)$, which represents the probability of selecting each intensity bin, is defined by $p(b_i) = n_i/N$.
- The output distribution $p(R)$, which represents the normalized area of each region r, is given by

$$p(r_j) = \sum_{i=1}^{N_b} p(b_i)p(r_j|b_i) = \frac{n_j}{N}$$

where n_j is the number of pixels of region r_j.

We adopt a greedy mutual-information-based algorithm[10] that splits the image in quasihomogeneous regions. The procedure takes the full image as the unique initial partition and progressively subdivides it (for example, in a binary space partition or quadtree) according to the maximum mutual information gain for each partitioning step. The algorithm generates a partitioning tree $T(\mathrm{I})$ for a given ratio of mutual information gain or a predefined number of regions (N_r is the number of tree leaves).

The mutual information gained in this decomposition process qualifies an image's capacity to be ordered or the feasibility of an observer decompositing it.

We can visualize this process from

$$H(B) = I(B,\hat{R}) + H(B|\hat{R})$$

where $\hat{R}$ is the random variable that represents the set of regions of the image and varies after each new partition. Information acquisition increases $I(B,\hat{R})$ (data processing inequality[4]) and decreases $H(B,\hat{R})$, producing an uncertainty reduction due to the regions' equalization. The maximum mutual information that we can achieve is $H(B)$.

We consider that the resulting tree captures the image's structure and hierarchy, and the mutual information gained in this decomposition process quantifies an image's capacity to be ordered or the feasibility of an observer decomposing it. Thus, varying the output from the single image until the lowest level (that is, the pixels) lets us study the information in the image's composition. The further down the regions we must go to achieve a given

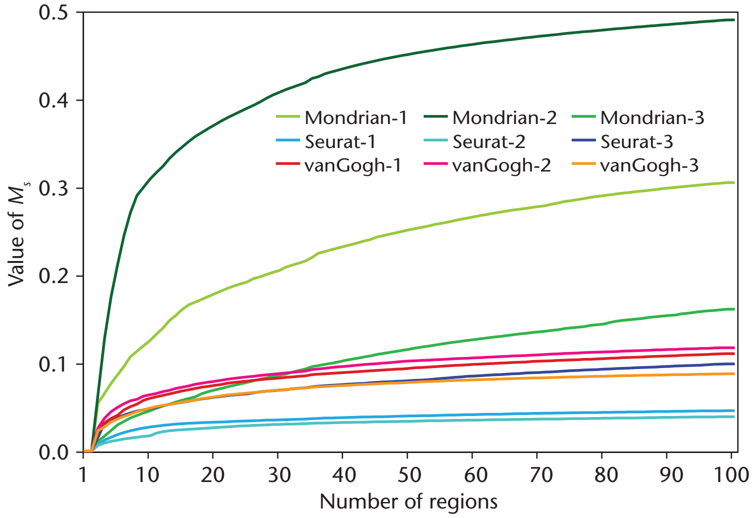
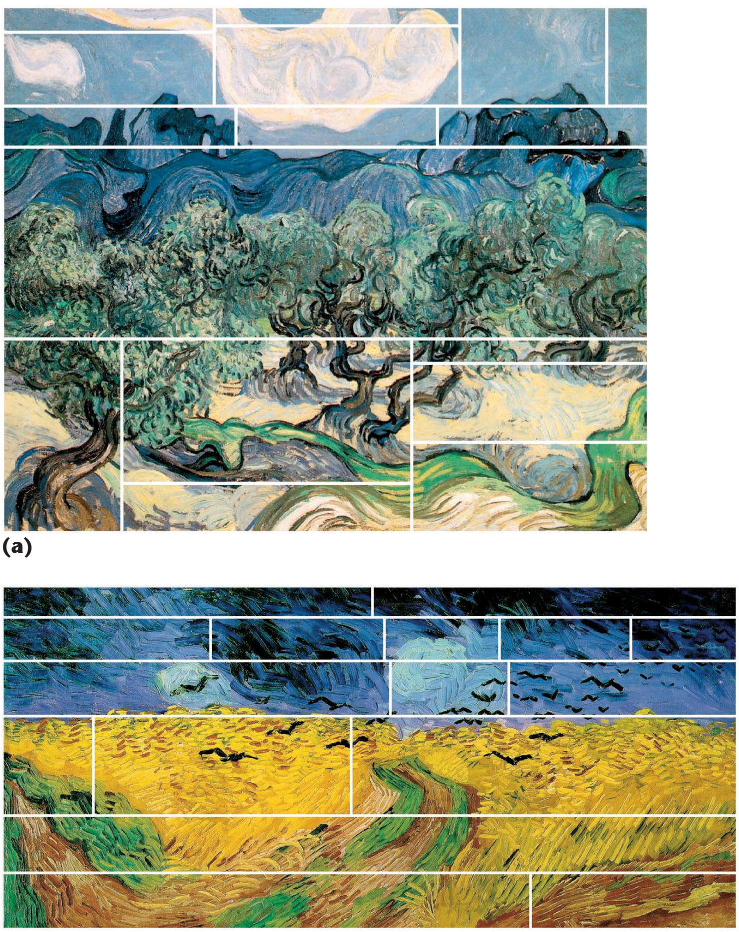


Figure 3. Evolution of ratio M_s for the set of paintings in Figure 1 (the first 100 splits).



(a)

(b)

Figure 4. Decompositions of (a) van Gogh-2 and (b) van Gogh-3 for $M_s(16) = 0.074$ and $M_s(16) = 0.057$, respectively (see Equation 6).

level of information, the more complex the image.

Similarly to Bense's communication channel between the repertoire and the final product, the channel we introduced can serve as the information (or communication) channel that expresses color distribution on a canvas. So, given an initial entropy or uncertainty of the image and a predefined level of resolution n, the evolution of the ratio

$$M_s(n) = \frac{I(B,\hat{R})}{H(B)} \tag{6}$$

represents the distribution process. Note that n ranges $1 \le n \le N_r^{\max} \le N_r$, where $N_r^{\max}$ is the minimum number of regions that provide all the image information (that is, $M_s(N_r^{\max}) = 1$).

Figure 3 shows the evolution of M_s building a binary space partitioning (BSP) for each painting in Figure 1. The capacity of extracting order from each painting coincides with the behavior expected by an observer. Note the grouping of the three different painting styles. Table 3 shows M_s values for n = 16 for the set of paintings. Although the partitioning reflects the geometry and randomness of Mondrian's and Seurat's paintings, respectively, it also finds the landscape elements in van Gogh's paintings (see Figure 4). Finally, Figure 5 shows a sequence of decompositions of van Gogh-1 obtained for several values of n, and only accounting for the luminance. Each region is painted with the average color corresponding to that region. With relatively few regions, the painting's composition is already visible (see Figures 5c and 5d), although the details aren't sufficiently represented.

We studied the image composition using an adaptive algorithm that partitions the image using a BSP structure driven by the maximum information gain at each partition. This algorithm shows us how the image's composition (macro-aesthetic description) appears clearly after relatively few partitions. On the contrary, the details or forms in the painting appear when we reach a refined mesh (microaesthetic description).

Our three compositional measures capture the spatial order in an image from an informational viewpoint. The first two measures, M_j and M_k, measure similarities between predefined regions using the information content (Shannon entropy) and the algorithmic complexity (Kolmogorov complexity), respectively. The third measure, M_s, based on shared information (mutual information), goes one step further by dynamically evolving as the structure is discovered.

Conclusion

Further work will explore the use of higher-order Shannon measures, such as entropy rate and excess entropy, which can help us to understand a painting's compositional aspects. Following Zurek's work,[8] we'll also analyze the artistic process from the viewpoint of a Maxwell's demon-type

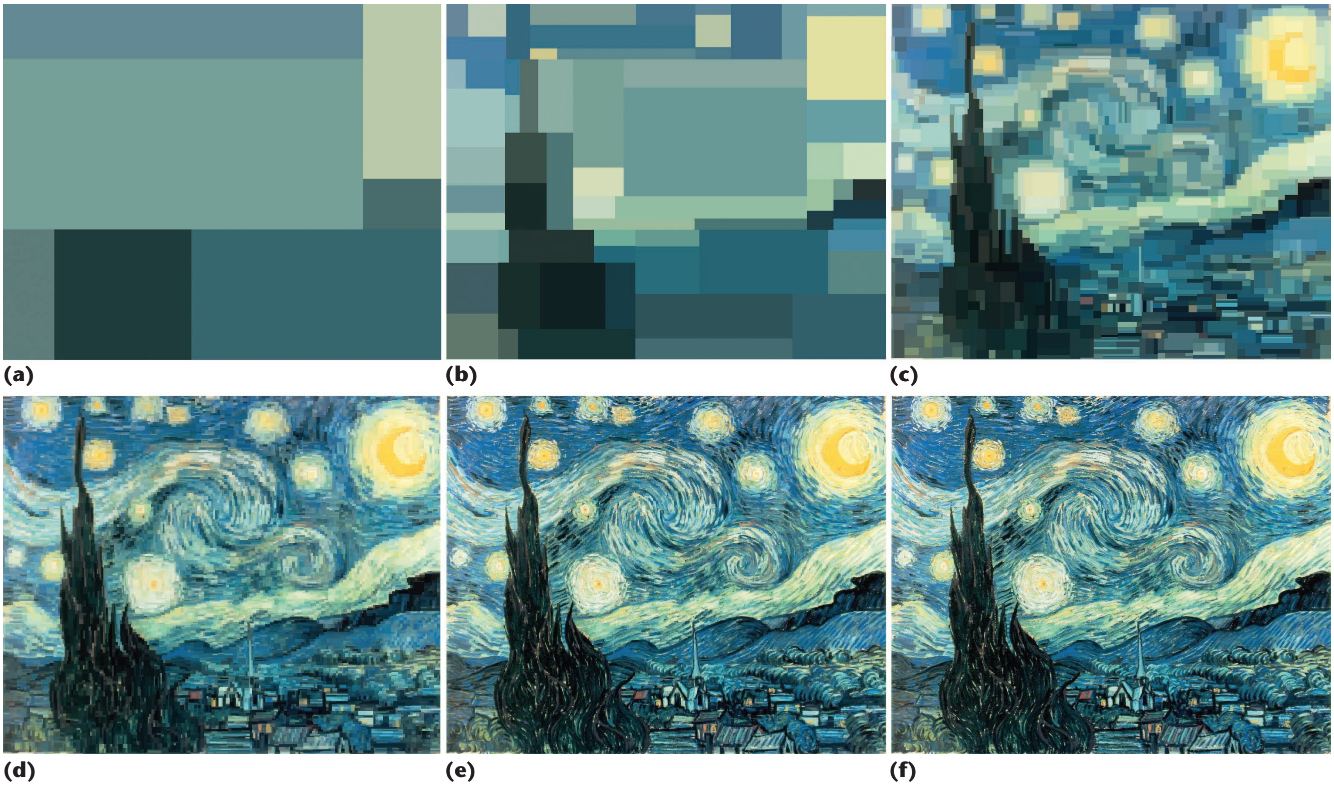


Figure 5. Evolution of an adaptive decomposition of van Gogh-1 using our partitioning algorithm. For each subfigure, we show the corresponding ($M_S(n)$, n): (a) (0.05, 7), (b) (0.1, 66), (c) (0.2, 1,574), (d) (0.4, 17,309), (e) (0.8, 246,573), and (f) (1.0, 789,235). We need a total of 789,235 regions to achieve the total information (that is, N_r^{max} = 789,235).

artist or observer, capable of selecting and classifying the information contained in an object.

We'll test our measures against a broader collection of paintings and other artwork, such as artistic photography. We'll conduct experiments within and across styles or painters. Our aim is to investigate the possibility of using these measures for classifying styles, or for distinguishing periods within a given artist's life. An interesting experiment will be to compare the automatic classification obtained by our measures with human experts' classification. We believe that we're on a promising track with a sound theoretical basis, which not only extends but will further develop Birkhoff's and Bense's aesthetics studies.

References

1. G.D. Birkhoff, *Aesthetic Measure*, Harvard Univ. Press, 1933.
2. M. Bense, *Einführung in die informationstheoretische Asthetik. Grundlegung und Anwendung in der Texttheorie* (Introduction to the Information-theoretical Aesthetics. Foundation and Application to the Text Theory), Rowohlt Taschenbuch Verlag, 1969.
3. A. Moles, *Information Theory and Esthetic Perception*, Univ. of Illinois Press, 1968.
4. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, John Wiley & Sons, 1991.
5. R. Scha and R. Bod, "Computationele Esthetica," (Computational Esthetics), *Informatie en Informatiebeleid*, vol. 11, no. 1, 1993, pp. 54-63; English translation available at http://iaaa.nl/rs/compestE.html.
6. J. Rigau, M. Feixas, and M. Sbert, "Conceptualizing Birkhoff's Aesthetic Measure Using Shannon Entropy and Kolmogorov Complexity," *Proc. Eurographics Workshop Computational Aesthetics in Graphics, Visualization and Imaging*, Eurographics Assoc., 2007, pp. 105-112.
7. M. Li et al., "The Similarity Metric," *IEEE Trans. Information Theory*, vol. 50, no. 12, 2004, pp. 3250-3264.
8. W.H. Zurek, "Algorithmic Randomness and Physical Entropy," *Physical Rev. A*, vol. 40, no. 8, 1989, pp. 4731-4751.
9. F. Nake, *Ästhetik als Informationsverarbeitung: Grundlagen und Anwendungen der Informatik im Bereich ästhetischer Produktion und Kritik* (Aesthetics as Data Processing: Bases and Applications of Computer Science in the Area of Aesthetic Production and Criticism), Springer, 1974.
10. J. Rigau, M. Feixas, and M. Sbert, "An Information Theoretic Framework for Image Segmentation," *Proc. IEEE Int'l Conf. Image Processing* (ICIP 04), vol. 2, IEEE Press, 2004, pp. 1193-1196.

Acknowledgments

The Spanish Ministry of Education and Science partly funded this work through grant number TIN2007-68066-C04-01.

Jaume Rigau *is an associate professor of computer science at the University of Girona, Spain. His research interests include the application of information theory to computer graphics and image processing. Rigau received a PhD in computer science from the Technical University of Catalonia. Contact him at rigau@ima.udg.edu.*

Miquel Feixas *is an associate professor of computer science at the University of Girona, Spain. His research interests include the application of information theory to computer graphics and image processing. Feixas received a PhD in computer science from the Technical University of Catalonia. Contact him at feixas@ima.udg.edu.*

Mateu Sbert *is a professor of computer science at the University of Girona, Spain. His research interests include application of Monte Carlo and information theory to computer graphics and image processing. Sbert received a PhD in computer science from the Technical University of Catalonia. Contact him at mateu@ima.udg.edu.*

For further information on this or any other computing topic, please visit our Digital Library at http://www.computer.org/csdl.