;

PAINTING THE BELL CURVE

BILL J RALPH

ABSTRACT. A new image statistic is introduced whose distribution is found to be nearly a Bell curve for many of the paintings of great artists. In support of this observation, the average skewness and excess kurtosis are shown to be close to zero for samples from the works of many painters from across the centuries. We find that not all images share these properties because the skewness sampled from photographs is generally far from zero. The analysis will consider multiple works by dozens of famous artists and contrast them with the results obtained for several different collections of photographs.

1. Introduction

The evidence we provide here suggests that the works of great painters often exhibit a statistical property that is not common in photographs. To detect this difference, we will define a new statistic in Section 7 called Flux and use its logarithm LFlux to measure colour variations near points in an image. When the LFlux is sampled at large numbers of points in certain images we find that its histogram closely approximates a Bell curve.

As a first example, consider the histogram shown in Figure 1 obtained by sampling the LFlux at a large number of points from Cezanne's painting "Quartier Four, Auvers-sur-Oise" circa 1873 . The solid line shows an ideal bell curve and the dots show the distribution of the LFlux. The fit is remarkably good well into the tails. The left tail is generated by smoother regions of the image and the right tail comes from points where there are sharp colour contrasts. For this painting, the LFlux statistic has a skewness of 0.0003 and an excess kurtosis of -0.0601 which are consistent with the zeroes we would expect from a normal variable. There is a transformation that changes a normal variable into a uniform one. After applying this conversion to the data, the histogram shown in Figure 1 is transformed into the one indicated

Date: June 30, 2020.
Key words and phrases. image statistics, Bell curve, visual art, photography, painting, aesthetics .

by the dots in Figure 3. The uniformity in Figure 3 suggests that the original variable was close to normal.

By way of contrast consider the histogram of the LFlux obtained for Warhol's "Campbells Soup I: Tomato 46" shown in Figure 2 which is significantly different from a Bell curve. The LFlux sample from this work had a skewness of 0.3549 and an excess kurtosis of 1.6102 which are significantly different than the zeros we would expect from a normal distribution. After applying the conversion from normal to uniform, we obtain the histogram for the LFlux indicated by the dots in Figure 4. This histogram is not uniform so the original variable was not normal. In this paper, we will see that many great artists have paintings with a skewness and excess kurtosis of zero as we would expect for a normal variable.

There has been a relatively small amount of research into the statistical analysis of fine art. Ruderman in [5] considers invariance to scale in natural images. The possibly fractal nature of Pollock's drip paintings has been explored in [6] and the history of painting has been examined through the lens of entropy in [7]. Our goal is to identify a new consistent pattern in the works of certain recognized artists.

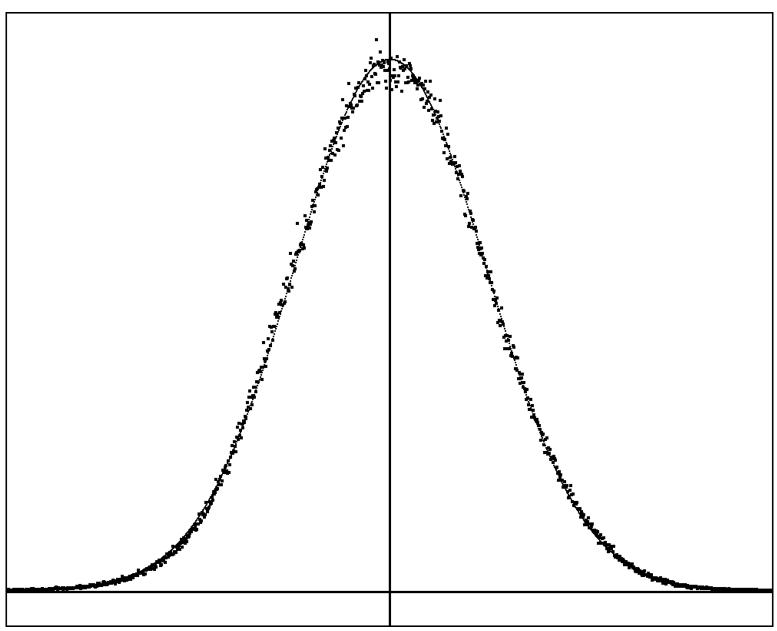

FIGURE 1. Transformed LFlux histogram for Cezanne's "Quartier Four, Auvers-sur-Oise"

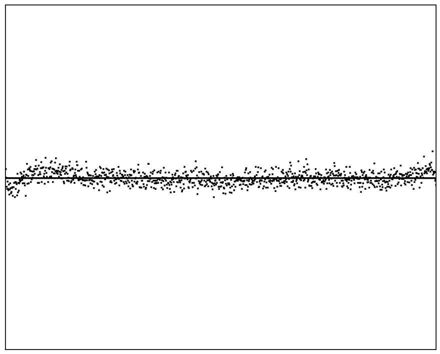

FIGURE 3. Transformed LFlux histogram for Cezanne's "Quartier Four, Auvers-sur-Oise".

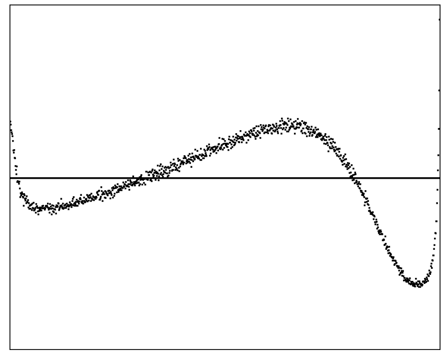

FIGURE 4. Transformed LFlux histogram for Warhol's "Campbells Soup I: Tomato 46"

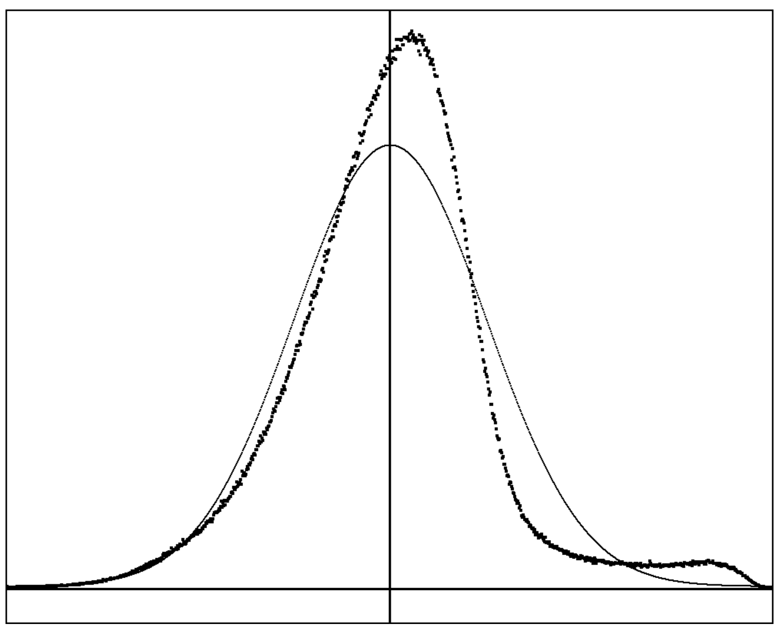

FIGURE 2. LFlux histogram for Warhol's "Campbells Soup I: Tomato 46"

2. Method And Data

Our perceptions of both colour and scale must be incorporated into any statistics designed to reflect our subjective experience of images.

We will consider colour first. The Lab colour model consists of a lightness component L and two chromatic components a and b. Lab takes into account our nonlinear perceptions of colour in order to provide a perceptually uniform metric on colour space. This metric forms the basis for the definition of Flux given in Section 7 and its uniformity gives us some confidence that Flux is measuring colour variations in a subjectively meaningful way.

In regard to scale, we observe that apart from resolution, images are not significantly distorted when viewed through quite a range of viewing distances. In order to eliminate considerations of scale, we will only consider image statistics that are scale invariant. The standard deviation, skewness and kurtosis of the LFlux all have this property as is easily seen from the definition of Flux. We will focus on just the skewness and excess kurtosis as a way of measuring how much an artist's work deviates from having a bell curve for its LFlux histogram.

Our samples are taken from the digitized paintings in [4] where each image contains roughly half a million to a million pixels. The first step in analyzing an image is to convert the RGB coordinates into Lab coordinates. The Flux is sampled at each pixel of an image and this data is used to construct a histogram of the LFlux and to estimate the skewness and kurtosis of that image. Table 1 and Table 2 show the average skewness and average excess Kurtosis of the LFlux obtained from several sample paintings for each artist and the numbers in brackets show their standard deviations.

We will limit the question of whether or not the LFlux has a normal distribution to simply checking that the sample data has both a mean of zero and an excess kurtosis of zero which suggests the use of the JarqueBera test statistic. We use a modified version of this statistic called Discrepancy which we define by $Discrepancy = S^2 + \frac{1}{4}(K-3)^2$ where S is the skewness and K is the kurtosis. The average Discrepancy for a selection of each artist's paintings is shown in the last column of the tables along with its standard deviation in brackets. From now on skewness and kurtosis will always refer to the skewness and excess kurtosis of the LFlux.

Artist	*Samp Size*	*Skewness*	*Excess Kurtosis*	*Discrepancy*
Avery	15	0.5699(0.3166)	0.6547(0.9485)	0.7354(0.7586)
Bacon	19	0.3515(0.5950)	0.2055(0.6583)	0.5721(0.5675)
Basquiat	15	0.2124(0.14089)	−0.6415(0.2470)	0.1807(0.0951)
Beckman	35	0.01811(0.2336)	−0.1622(0.3084)	0.0830(0.1210)
Bonnard	17	−0.1904(0.2204)	0.0892(0.2063)	0.0940(0.0842)
Botticelli	8	−0.0610(0.1790)	−0.1107(0.2604)	0.04963(0.0458)
Bruegel	16	−0.0915(0.1385)	−0.2289(0.2081)	0.04960(0.0408)
Caravaggio	12	0.2323(0.2451)	−0.3116(0.4217)	0.1741(0.1129)
Cassatt	23	0.1126(0.2072)	0.0283(0.2158)	0.0651(0.0692)
Cezanne	92	−0.0945(0.2395)	0.1129(0.2393)	0.0830(0.1139)
Chagall	17	0.0984(0.2194)	−0.1588(0.2054)	0.0712(0.0875)
Constable	22	−0.0228(0.1772)	−0.2098(0.2947)	0.0622(0.0525)
Corot	16	−0.1694(0.1891)	−0.1444(0.2590)	0.0831(0.0969)
Courbet	26	0.0550(0.1919)	0.0537(0.1974)	0.0485(0.0606)
David	21	0.2196(0.2121)	0.0280(0.3428)	0.1193(0.1406)
De Chirico	14	0.1421(0.1921)	0.0316(0.2279)	0.0668(0.0719)
Degas	39	0.0022(0.1710)	0.1195(0.1670)	0.0389(0.0432)
Delacroix	23	−0.1199(0.2173)	−0.0681(0.2389)	0.0743(0.1063)
El Greco	32	−0.0520(0.4384)	0.1249(0.6145)	0.2843(1.1097)
Gaugin	37	−0.0590(0.2029)	0.0638(0.2082)	0.0550(0.0536)
Gris	16	0.0525(0.4365)	−0.0406(0.4037)	0.2200(0.1986)
Hockney	21	0.4185(0.3621)	0.0698(0.6396)	0.3986(0.4913)

Table 1. The mean and standard deviation of three different statistics sampled from paintings by different artists.

Artist	*Samp Size*	*Skewness*	*Excess Kurtosis*	*Discrepancy*
Hopper	32	$-0.0541(0.2602)$	$-0.0242(0.3103)$	$0.0920(0.1141)$
Inness	12	$-0.0459(0.1008)$	$0.1588(0.1208)$	$0.0211(0.0143)$
Kahlo	19	$0.2686(0.2554)$	$-0.0075(0.4007)$	$0.1720(0.2025)$
Kandinski	16	$0.2499(0.2591)$	$0.1321(0.1142)$	$0.1015(0.1787)$
Kiefer	10	$-0.3874(0.1396)$	$0.02146(0.2364)$	$0.1612(0.0997)$
Lichtenstein	17	$0.1485(0.3725)$	$-0.6668(0.4581)$	$0.3132(0.1641)$
Manet	19	$0.2681(0.2787)$	$0.1101(0.5157)$	$0.2115(0.3742)$
Matisse	40	$0.3633(0.4755)$	$0.1129(0.6738)$	$0.4663(0.8199)$
Monet	32	$-0.1409(0.3248)$	$-0.0978(0.4174)$	$0.1666(0.2200)$
Morisot	8	$-0.0810(0.2746)$	$-0.0136(0.3634)$	$0.10148(0.1787)$
Picasso	56	$0.1043(0.3133)$	$0.0008(0.3418)$	$0.1359(0.2015)$
Raphael	17	$0.1416(0.2774)$	$-0.0213(0.2313)$	$0.1052(0.1225)$
Rembrandt	34	$-0.0037(0.1887)$	$-0.1177(0.3426)$	$0.0665(0.0561)$
Renoir	29	$-0.1328(0.2171)$	$0.1243(0.2329)$	$0.0801(0.1091)$
Rubens	34	$-0.0227(0.2163)$	$-0.1273(0.2919)$	$0.0707(0.0781)$
Sargent	20	$0.2645(0.2040)$	$-0.01145(0.3426)$	$0.1374(0.1187)$
Titian	33	$0.0254(0.2139)$	$-0.0868(0.2698)$	$0.0645(0.07830)$
Turner	22	$0.0549(0.2179)$	$0.0811(0.4370)$	$0.0955(0.1743)$
Van Gogh	52	$-0.1755(0.2199)$	$-0.07442(0.3320)$	$0.1066(0.0998)$
Velazquez	33	$0.1950(0.2840)$	$-0.0630(0.4285)$	$0.1617(0.4635)$
Vermeer	20	$0.1303(0.3791)$	$-0.1449(0.3182)$	$0.1828(0.2672)$

Table 2. The mean and standard deviation of three different statistics sampled from paintings by different artists.

3. Overview Of The Data

The Lab colour model deals with the nonlinear nature of vision so the Flux has already taken into account that our perception of colour is somewhat logarithmic. The only reason to take the logarithm a second time is to test whether the LFlux is normally distributed or equivalently that the Flux has a lognormal distribution. Since LFlux is a logarithm, small differences in the values of its statistics can correspond to large differences in the appearances of images. For example, we attribute some of the visual differences in the images used to generate Figure 1 and Figure 2 to their very different skewnesses of approximately 0.0 and 0.67 respectively. A look through the tables shows that no artist has an average skewness greater than or equal to 0.67 and most artists are in the vicinity of 0.

If we average the statistics in the two tables, we find the average skewness is 0.0670 and the average excess kurtosis is -0.0283 which are remarkably close to 0. The skewness drops to 0.03 if we remove the outliers with an absolute value greater than 0.3. The average Discrepancy for each artist is shown in the last column of the tables and is a measure of the extent to which the artist's paintings fail to have both a skewness and excess kurtosis equal to zero. Amongst painters working before 1700, Rembrandt, Titian and Rubens stand out for the exceptionally low values of the Discrepancy. Table 3 shows averages of the values given in Table 1 and Table 2 for different time periods. After 1700 the kurtosis remains close to zero and before 1880 the skewness is also close to zero. After 1880, the skewness moves to the right as artists work becomes a bit more graphical perhaps influenced by photography and a desire for greater clarity.

Time Period	*Skewness*	*Excess Kurtosis*	*Discrepancy*
Before 1700	0.0493	-1.0873	0.1209
1700 - 1800	-0.0075	-0.0626	0.0869
1800 - 1880	0.0343	0.0544	0.1189
After 1880	0.1494	-0.0552	0.2463

TABLE 3. Averages of the LFlux statistics for artists born in different time periods.

Figure 5 shows kurtosis plotted against skewness for all the entries in the tables and we observe some clustering around $(0, 0)$. Some of the outliers on this graph have been numbered and correspond to the following artists:

(1) Lichtenstein
(2) Basquiat
(3) Caravaggio
(4) Avery
(5) Kiefer

Artistic traditions might be part of what keeps many artists work close to the origin in Figure 5. But many painters are driven to seek new modes of expression which might be the reason artists, like the outliers above, staked their claim on some of the empty territory in Figure 5. Fortunately, there is still a lot of room for future artists to explore.

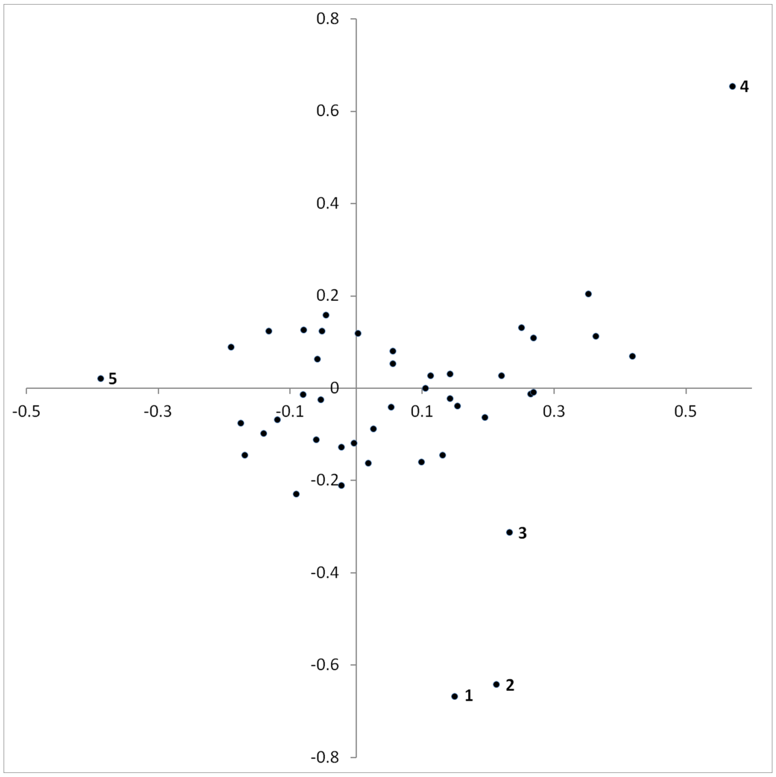

Figure 5. Average excess kurtosis plotted against average skewness for 44 well known artists

4. Interpreting The Skewness And Kurtosis

If the LFlux is skewed to the right, as for example in Bacon's "Self Portrait" where it is 1.07, then we see well defined foregrounds against relatively simple background textures. For paintings skewed to the left, like Kiefer's "Nuremberg" where it is -0.52, the images appear to be less well defined with thicker and more complex textures. The explanation for these variations in appearance lies in the way the histogram captures the large and small colour contrasts in the image. If the LFlux histogram is skewed to the right, then there are more places with small colour contrasts and our eyes are drawn to the relatively smaller number of sharp contrasts. This tight focus on a small number of points results in an overall impression of greater clarity. On the other hand, if the LFLux is skewed to the left, then the relatively large number of places with sharp colour contrasts creates a busy image that we might interpret as being more textured. In short, if an artist emphasizes clarity over texture then the skewness will be positive and if the reverse is true then the skewness will be negative.

The impressionist painters listed in the tables all have an average skewness that is close to zero or slightly negative due to their greater emphasis on texture. Portrait painters like David and Sargent generally position figures against quiet backgrounds so we expect their work to be skewed to the right which is reflected in their table entries.

After the impressionists, there was a major shift in artists work toward positive skewness as shown by the skewness timeline in Figure 6 which plots the average skewness of an artist's works against their date of birth. This graph reflects the tendency of some artists in the 20th century to create work with greater clarity. Examples of such works are Matisse's "Icarus" from 1947 with a skewness of 1.5733 and Kandinsky's "Transverse Line" from 1923 with a skewness of 0.6362. These works are clearly more graphically oriented than the works of the impressionist period.

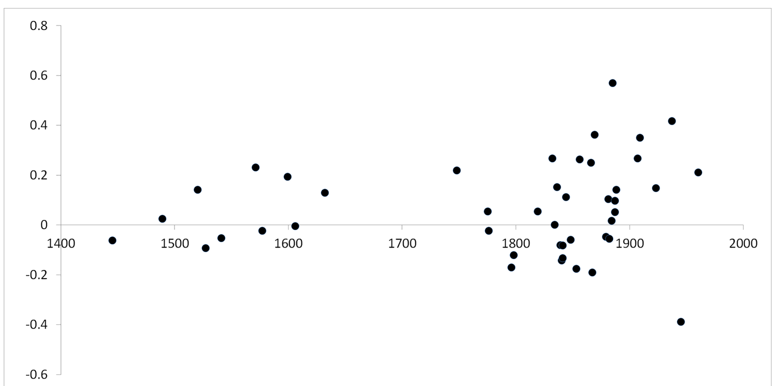

FIGURE 6. Average skewness of an artist's work plotted against their date of birth.

Even in mainstream statistics there has been a certain amount of controversy concerning the interpretation of kurtosis. Whatever the mathematical interpretation, artists from 1700 onward have stayed close to a kurtosis of 0 on average as indicated in Table 3. Extreme examples are Avery's "Black Sea" with a kurtosis of 2.6877 and Basquiat's "Logo" with a Kurtosis of -1.2366. In the "Black Sea", we see that points where the colours change abruptly are quite rare whereas in "Logo" there are regions that are flat and regions that are filled with very fine detail comprised of abrupt changes in colour. Despite these extremes, the tables show that the majority of artists stay close to a kurtosis of 0 on average.

5. How Are Paintings Different From Photographs

We will compare the artists listed in the tables to the photographs in the five collections listed in Table 4. In this last table, the Nature and Urban collections are from [1], the Radom collection is from [2], the Interiors collection is discussed in [3] and the Florida photographs were taken on vacation by the author. Discrepancy is our measure of how much the LFlux histogram differs from a bell curve. Eighty percent of the artists in the tables have discrepancies less than 0.2 but each of the five collections of photographs has an average discrepancy greater than 0.2. Moreover, the standard deviation of the discrepancy for most artists is small whereas the standard deviations for four of the photograph collections are very large. In other words, the works of many of the artists in the tables are fairly close to having a bell curve as their LFlux histogram but most photographs are very far from having bell curves. One way to compare artists' works and photographs is to plot them both on a kurtosis-skewness diagram that we now consider.

The circular dots in Figure 7 show the kurtosis plotted against the skewness for 330 pictures taken by the author on a Florida trip as well as the parabola of best fit. The distribution of these dots is typical of what is found for collections of photographs. By way of comparison, the black squares plot the values listed for each artist in the tables. The artists are fairly tightly clustered around $(0, 0)$ while the photographs follow a wide range that mostly avoids $(0, 0)$. This graph is a bit misleading since we are using averages for the artists works and not for the photographs. However, artists like Rembrandt, Titian, Reubens and Turner would fit quite tightly around $(0, 0)$ in this graph as can be seen by the discrepancies given in the tables.

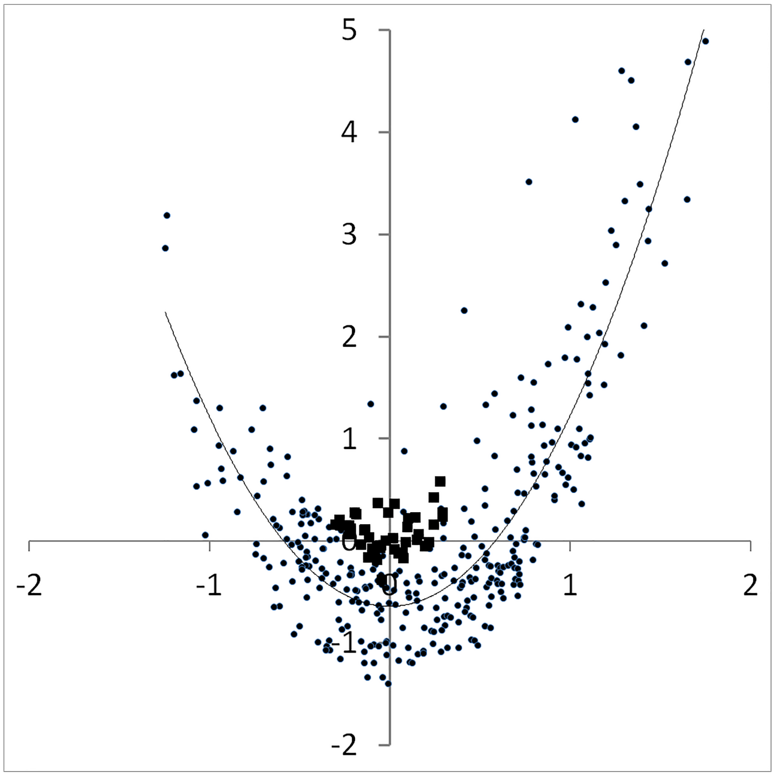

FIGURE 7. Kurtosis plotted against skewness where the round dots are Florida vacation pictures and the squares are the values listed for each artist in the tables.

Collection	*Samp Size*	*Skewness*	*Excess Kurtosis*	*Discrepancy*
Nature	300	−0.0748(0.4483)	−0.1495(0.5813)	0.4106(1.9203)
Urban	300	−0.0495(0.3307)	0.3297(0.6296)	0.2373(0.2738)
Random	64346	0.2285(0.5028)	0.0252(1.0376)	0.5743(2.4599)
Interiors	5230	0.6324(0.3774)	0.3084(0.9215)	0.7784(1.0336)
Florida	330	0.2045(0.6249)	0.1607(1.1324)	0.7572(1.2288)

TABLE 4. The mean and standard deviation of three different statistics sampled from different collections of photographs.

6. DISCUSSION

The skewness and kurtosis seen in the works of several well known artists show a large range which is not surprising given that art is such a highly personal and creative process in which artists are constantly searching for novel visual territory. But it is surprising to find that many of our most celebrated artists are creating works with a skewness and kurtosis close to zero and with histograms close to the bell curve shown in Figure 1. In order to guarantee normality, it is not sufficient for an image to have a skewness and kurtosis of zero. However, in all of our tests the histograms were very smooth and in only one case was it bimodal which leads us to believe that the central underlying distribution that artists are either working with or playing against is the normal distribution.

Artists like Bacon, Basquiat, Caravaggio and Hockney have rejected a skewness of zero in favour of works of greater clarity. Artists like Kiefer have taken the skewness in the opposite direction and created works that are thickly textured. Figure 6 suggests that having a skewness of zero may have been a component of the aesthetics of many artists that was gradually abandoned as art moved into the twentieth century. The skewness is still a useful statistic because it allows us to quantitatively track the movement of art away from what might be called traditional values.

Based on our analysis, the central question we have to ask is why the processes of some artists lead to the LFlux histogram being a bell curve? A clue might lie in the results from the Nature photographs summarized in Table 4. The average skewness and kurtosis of these photographs is fairly close to zero but the discrepancy is not only large but varies across a very wide range of values. Is it possible that artists are averaging their daily viewing experiences and creating paintings that match this average? Is it also possible that these paintings are aesthetically pleasing to us simply because in a single work we are experiencing the average of the enormous number of visual experiences that we have in our daily lives?

In future work, we will consider the coefficient of variation of the Flux which is the ratio of its standard deviation to its mean. This number is scale invariant and varies with the density of detail in an image. Information about the painting technique of an artist is contained in the mean and standard deviation of the Flux which are not scale invariant statistics and therefore require us to include the dimensions of the image in our calculations. These two statistics give us information about an artist's technique which will be discussed elsewhere.

7. The Definition Of Flux

We assume that the colour coordinates of the points in an image are given in Lab coordinates. Let $C : I \to R^3$ be the function on the set of pixels I in the image that assigns to each pixel (i, j) its colour coordinates $(L(i, j), A(i, j), B(i, j))$. In this section, we introduce a function $Flux : I \to (0, \infty)$ designed to detect local colour variations which is similar to the norm of the gradient of C. The function $Flux$ will be our primary tool for extracting information from a digital image. The norm of the gradient of a function $f : R^2 \to R$ is given by $|| \bigtriangledown (f)||$ $= \sqrt{(\frac{\partial f}{\partial x})^2 + (\frac{\partial f}{\partial y})^2}$ but for a digital image it will be more useful to use the easily verified fact that

$$(1) \qquad \begin{aligned} || \bigtriangledown (f)|| = \lim_{h \to 0} \sqrt{(}(f(x+h, y+h) - f(x, y))^2 \\ +(f(x, y+h) - f(x+h, y))^2)/(2h^2)) \end{aligned}$$

.

In order to extend this defintion to the function f, we first rewrite it in terms of the usual metric $d(a, b) = |a - b|$ to obtain

$$(2) \qquad \begin{aligned} || \bigtriangledown (f)|| = \lim_{h \to 0} \sqrt{(}d(f(x+h, y+h), f(x, y))^2 \\ +d(f(x, y+h), f(x+h, y))^2)/(2h^2)) \end{aligned}$$

Now replace d in equation 2 by the usual metric on R^3, set $h = 1$ and replace f by C to obtain the following explicit definition of $Flux$.

(3)

$$\begin{aligned} Flux(i, j) = \sqrt{(}(L(i+1, j+1) - L(i, j))^2 + (L(i, j+1) - L(i+1, j))^2)/2 \\ +((A(i+1, j+1) - A(i, j))^2 + (A(i, j+1) - A(i+1, j))^2)/2 \\ +((B(i+1, j+1) - B(i, j))^2 + (B(i, j+1) - B(i+1, j))^2)/2) \end{aligned}$$

Our choice of $h = 1$ is natural for an image comprised of pixels but in making this choice we are throwing away information about the scale of the image. In this paper, we only made use of the skewness and kurtosis of the logarithm of the Flux which are not affected by the value of h. However, if h is properly determined by the number of pixels and the size of the image, then the mean and standard deviation of the numbers $Flux(i, j)$ become meaningful descriptions of an artists technique that will be studied elsewhere.

Finally, we want to allow for the fact that humans have difficulty distinguishing colours whose distance apart is less than 1 in the Lab metric. For this reason, at each pixel (i, j) we will replace its colour values (L, a, b) by a triple $(L+s, a+t, b+u)$ where s, t and u are chosen independently and uniformly at random from the interval $[-0.5, 0.5]$.

References

[1] Artificial Intelligence Lab, Vrije Universiteit Brussel, Data Set At http://arti.vub.ac.be/research/colour/data/imagesets.zip

[2] Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li Jia-Li, David Ayman Shamma, Michael Bernstein, Li Fei-Fei

[3] Sean Bell, Kavita Bala, Noah Snavely. "Intrinsic Images in the Wild". ACM Transactions on Graphics (SIGGRAPH 2014). http://intrinsic.cs.cornell.edu.

[4] Harden, M., Mark Harden's Artchive, Online At https://www.artchive.com/

[5] Rudermen, D., The Statistics Of Natural Images, Network: Computation in Neural Systems 5 (1994) 517-548.

[6] Taylor, R., Fractal analysis of Pollocks drip paintings, Nature, 422, 399 (1999)

[7] Higor Y. D. Sigaki, Matja Perc, Haroldo V. Ribeiro, History of art paintings through the lens of entropy and complexity, Proceedings of the National Academy of Sciences Sep 2018, 115 (37)

Mathematics Department, Brock University, St. Catharines, Ontario, Canada L2S 3A1, Email: bralph@brocku.ca, Phone: (905) 931-3551