



OPEN

# Large-Scale Quantitative Analysis of Painting Arts

SUBJECT AREAS:  
SCIENTIFIC DATA  
STATISTICSDaniel Kim<sup>1</sup>, Seung-Woo Son<sup>2</sup> & Hawoong Jeong<sup>3,4</sup>

<sup>1</sup>Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea, <sup>2</sup>Department of Applied Physics, Hanyang University, Ansan 426-791, Korea, <sup>3</sup>Department of Physics and Institute for the BioCentury, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea, <sup>4</sup>Asia Pacific Center for Theoretical Physics, Pohang 790-784, Korea.

Received  
13 January 2014Accepted  
24 September 2014Published  
11 December 2014

Correspondence and requests for materials should be addressed to S.S. (sonswoo@hanyang.ac.kr) or H.J. (hjeong@kaist.edu)

Scientists have made efforts to understand the beauty of painting art in their own languages. As digital image acquisition of painting arts has made rapid progress, researchers have come to a point where it is possible to perform statistical analysis of a large-scale database of artistic paints to make a bridge between art and science. Using digital image processing techniques, we investigate three quantitative measures of images – the usage of individual colors, the variety of colors, and the roughness of the brightness. We found a difference in color usage between classical paintings and photographs, and a significantly low color variety of the medieval period. Interestingly, moreover, the increment of roughness exponent as painting techniques such as *chiaroscuro* and *sfumato* have advanced is consistent with historical circumstances.

Humans have expressed physical experiences and abstract ideas in artistic paintings such as cave paintings, frescos in cathedrals, and even graffiti on city walls. Such paintings, to convey intended messages, consist of three fundamental building blocks: points, lines, and planes. Recent studies have shed light on interesting mathematical patterns between these building blocks in paintings.

Artistic styles were analyzed through various statistical techniques such as fractal analysis<sup>1</sup>, the wavelet-based technique<sup>2</sup>, the multi-resolution hidden Markov method<sup>3</sup>, the Fisher kernel based approach<sup>4</sup>, and the sparse coding model<sup>5,6</sup>. Recently, these methods have also been applied to other cultural heritages such as literature<sup>7–10</sup> and music<sup>11–14</sup>. Such quantitative analysis is called “stylometry,” which originates from literature analysis to identify characteristic literary style<sup>9</sup>.

In this study, we add a new dimension to the body of stylometry studies by analyzing a large-scale database of artistic paintings. With digital image processing techniques we quantify the change in variety of painted colors and their spatial structures over ten historical periods of western paintings – *medieval, early renaissance, northern renaissance, high renaissance, mannerism, baroque, rococo, neoclassicism, romanticism, and realism* – starting from the 11th century to the mid-19th century. Digital images of the paintings were obtained from the Web Gallery of Art<sup>15</sup>, which is a searchable database for European paintings and sculptures consisting of over 29,000 pieces ranging from the years 1000 to 1850. Most of the identifiable images contain information of schools, periods, and artists, and are good quality in resolution to apply statistical analysis.

Here we focus on the following three quantities – the usage of each color, variety of painted colors, and the roughness of the brightness of images. First, we count how often a certain color appears in a painting for each period. From the frequency histogram, we find a clear difference between classical paintings and photographs. Next, we measure a fractal dimension of painted colors for each period in a color space, which is analogically considered to reflect the color ‘palette’ of that period. Interestingly, the fractal dimension of the medieval period is lower than that of other periods. The detailed results and our inference are discussed in this section. Last, we consider how rough or smooth an image is in the sense of its brightness. In order to quantify roughness of brightness, a well-known roughness exponent measurement in statistical physics is applied. We find that the roughness exponent increases gradually over the 10 periods, which is consistent with the historical circumstances like the birth of the new painting techniques such as *chiaroscuro* and *sfumato*<sup>16–17</sup> (*Chiaroscuro* and *sfumato* are major painting techniques developed and widely used during the Renaissance period. Literally, the compound word *chiaroscuro* is formed from the Italian words *chiaro* (light) and *oscuro* (dark), which refers to an artistic technique to delineate tonal contrasts and voluminous objects with a dramatic use of light. Precursors of *chiaroscuro* are Leonardo da Vinci (1452–1519) and Michelangelo Merisi da Caravaggio (1571–1610), and Rembrandt van Rijn (1606–1669) is a representative artist well-known for his use of *chiaroscuro*. The Italian word *sfumato* is derived from the Italian term *fumo* which literally means “smoke”. Leonardo da Vinci mentioned



*sfumato* as a blending of colors without lines or borders, in the manner of smoke or beyond the focus plane. In other words, *sfumato* is a painting technique to express gradual fade-out between object and background avoiding harsh outlines.). Analyzing these three properties, we propose new approaches to quantitatively analyze a large scale database of paintings. Applying our method to the controversial Jackson Pollock's drip paintings, it is possible to infer that his drip paintings are quite different from works of other painters.

## Results

**Chromo-spectroscopy.** First we investigate how many different kinds of color appear in a painting, and how often a certain color is painted, which is similar to Zipf's plot for word frequencies in literature<sup>18</sup>. It is named as "chromo-spectroscopy." A color is considered to be like a word for a painter. As an example of chromo-spectroscopy, Fig. 1a displays the fraction of each color used in a painting in descending rank order. If each color is chosen from a palette uniformly at random, the frequency of each color would follow a binomial distribution for a random process (see more detail in the supplement), and its rank plot would show an inverse of its cumulative, i.e., the regularized incomplete beta function<sup>19</sup>. This is because the rank plot is the inverse of its cumulative density function (see black dots in Fig. 1a). However, interestingly, the rank-ordered color-usage distribution (RCD) shows a long tail distribution, which is different from the inverse function of the regularized incomplete beta function (see Fig. 1a).

Figure 1b shows RCDs for 10 periods of European art history and photographs. The RCD of a period represents how many colors are used and how often a specific color appears during the period. All periods of painting show a universal distribution curve, but the rank of each color for each period is rather different. The RCD of photographs is similar to that of paintings at the beginning of a power-law part but the exponential tail deviates significantly from paintings, as shown in Fig. 1b. In order to clarify the difference of the tail section of RCDs between paintings and photographs, we analyze RCDs of images of photographs after applying several painting filters from popular software. There are clear changes in the tail of the distribution when only the oil painting filter is applied. An oil painting filter usually consists of two parameters – range and level – which are related to the size of an art paint brush and smearing intensity. It seems these two parameters influence the shape of the exponential tail of the RCD. Another interesting fact is that there is no clear

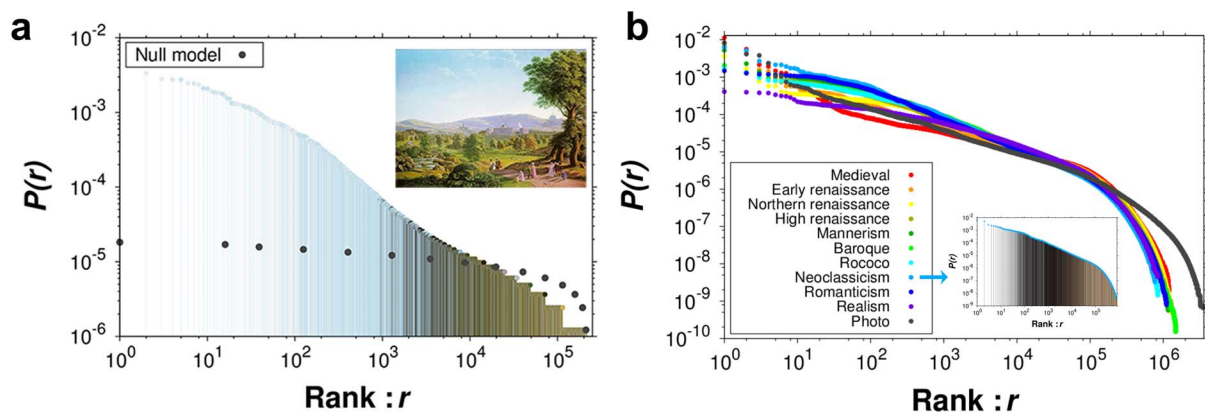
difference between RCDs of photographs and hyper-realism paintings, which are extremely finely drawn with microscope and are hard to distinguish from photographs with unaided eyes (see Figure S4b in the supplement). This suggests that paintings are only quantitatively distinguished from photographs by the tail section of the RCD. The tail of RCD represents frequency of noisy colors or a level of details in the image.

**Fractal pattern and color palette.** RCDs for all periods of paintings show quite universal distribution curves. However, the most commonly painted color is different for each period. To characterize the variety of colors more quantitatively, while ignoring its individual frequency, we investigate the fractal pattern of the painted color in the RGB color space for each period.

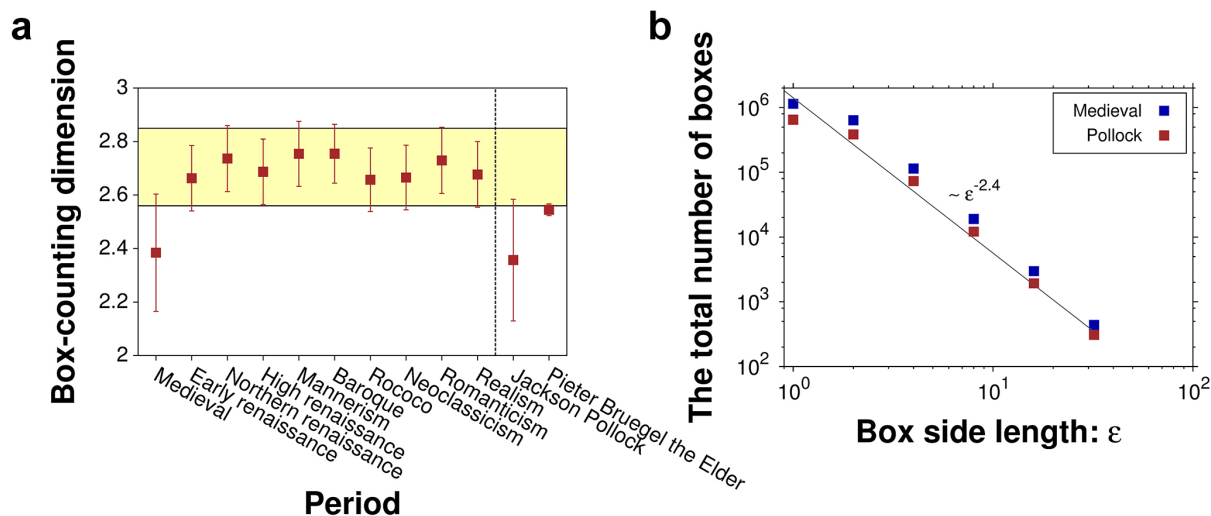
To examine the fractal characteristics of painted colors for each period, we measure the box-counting dimension<sup>20</sup> of the paintings in the RGB color space and compare them with two iconoclastic artists: Pieter Bruegel the Elder and Jackson Pollock. Each color used in the painting is plotted on a point in the RGB color space. Based on the definition of the box-counting dimension, we iteratively change the length of box  $\varepsilon$  from  $\varepsilon = 1$  to  $\varepsilon = 32$ , and count the number of non-empty boxes. A non-empty box indicates that corresponding colors within the box are used in the painting at least once. If the distribution of colors in the color space is homogeneous, the box counting dimension is 3. In other words, if the box counting dimension is less than 3, the distributions in the color space is heterogeneous and fractal, which means some axes are preferred or the distribution is composed of a preferred color scheme in the color space. In this sense, measuring the box-counting dimension quantifies the spatial uniformity or fractality of painted colors for each artistic period.

Figure 2a shows that the box-counting dimensions of paintings from the 10 historic periods are in the range between 2.6 and 2.8 except for the medieval period. As Fig. 2b shows, only the box-counting dimension of the medieval period is close to that of Jackson Pollock's drip paintings (below 2.4), where he used limited colors intentionally. In addition, the box counting dimension for the paintings of Pieter Bruegel the Elder is approximately 2.55. A low box-counting dimension represents that there is a strong preference in a small number of selected colors in the medieval age. That is, the color palette in the medieval age is significantly different from the other periods.

One can find the reason why the box counting dimensions for the medieval age and Jackson Pollock are different from others in the



**Figure 1 | Rank-ordered color-usage distributions for an image and periods.** (a) Fraction distribution of each color in a descending rank order for the artwork of German painter Johann Erdmann Hummel (1769-1852), “*Schloss Wilhelmshöhe with the Habichtswald*” (This image is out of copyright.). The horizontal axis indicates the rank of a color in frequency and the vertical axis denotes the proportion of a color in an image. The most (least) used color is located at the leftmost (rightmost) position on the horizontal axis. The black dots represent color choices from the same palette uniformly at random. (b) Rank-ordered color-usage distributions (RCDs) of the 10 periods and photographs. Note that the distribution of photographs clearly shows a different tail. Inset: RCD for the neoclassicism period. The displayed color corresponds to its rank. Note that the fraction is normalized by the image size and the number of paintings in each period.



**Figure 2 | Box-counting dimension and its tendency.** (a) The results of box-counting dimension over the 10 artistic periods display a significant difference of the medieval period from the other periods. Error bars indicate the standard deviation. (b) The number of boxes to cover the color space versus box size. The fractal dimension in the color space of Jackson Pollock's drip paintings is measured around 2.35, similar to that of medieval paintings (see also Figure S5 in the supplement), but dissimilar to that of another iconoclastic artist Pieter Bruegel the Elder.

historical facts. First, specific rare pigments were preferred for political purposes and religious reasons in the medieval age despite their expensive cost. Second, no technique of physical mixing between different pure colors was used in that period due to the tendency to emphasize the purity of colors and materials themselves. Artists recoated on a colored canvas to represent various colors in the middle age. The drip paintings of Jackson Pollock are also formed from recoating each single color dripping pattern on other layers, and the number of used colors is smaller than other western paintings before 20<sup>th</sup> century. Furthermore, oil colors and color mixing techniques were not fully developed until the Renaissance age. The introduction of new expression tools, like pastels and fingers, and painting techniques, such as *chiaroscuro* and *sfumato*, made much more colorful and natural expressions possible after the Renaissance period<sup>21</sup>. The difference of fractal dimensions between the medieval and other periods quantitatively may quantitatively reflect the historical facts and the painting technical difference in art history.

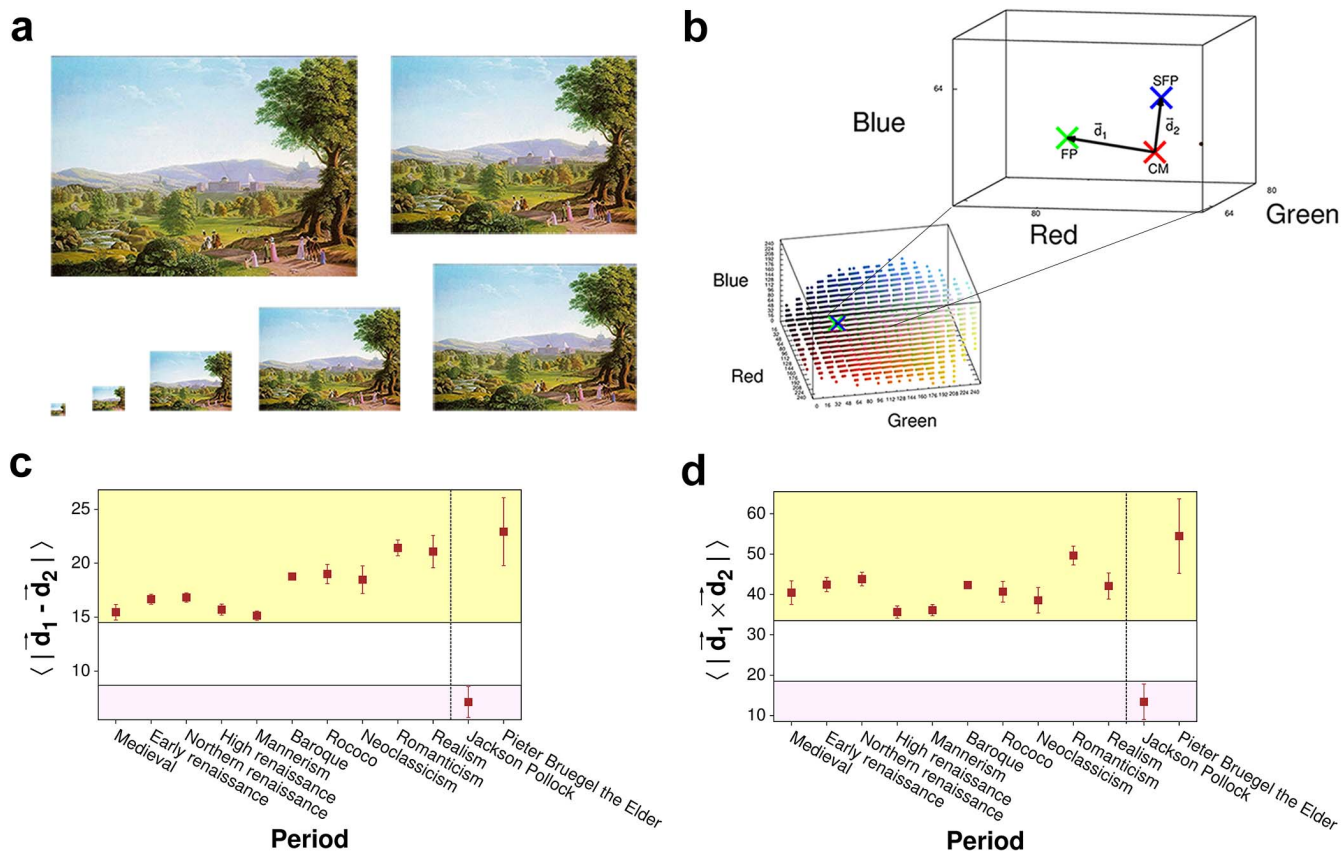
**Spatial renormalization and fixed point analysis.** In the RGB color space, each painting has its own set of scattered color pixels. In order to analyze the characteristics of color usages, considering the variety of color in the paintings, we define three representative points in the RGB color space. First, center of usage frequency in the color space may be compared to center of mass in physics. One can calculate center of usage frequency (CM) in the color space with the usage information and spatial position of colors such as the center of mass of physical objects. Second, iteratively resizing a painting is necessary to get the fixed point of the painting borrowed from real space renormalization concept in physics. Repeatedly resizing a painting, a painting eventually becomes one pixel. That is the fixed point of the painting (FP). The third fixed point of the randomized painting (SFP) is the same as mentioned in the second one except for shuffling the pixels of the painting. If the spatial information of the scattered color is irrelevant, FP and SFP would not be significantly different. Note that center of mass point of a shuffled image (SCM) is the same as the original CM. Then, two vectors  $d_1$  ( $d_2$ ) pointing from CM to FP (SFP) can be compared to quantify the randomness of the spatial arrangement of the colors in paintings. If  $d_1$  and  $d_2$  are similar, the used colors in a painting are not diverse or the spatial arrangement of the colors in a painting is close to random. Figure 3c suggests that the color arrangement of Jackson Pollock's drip paintings is quite different from other paintings, showing that

Pollock's art work is quite random, especially in the spatial arrangement of colors. On the other hand, the two fixed points of Pieter Bruegel the Elder's paintings are far away each other.

**Surface roughness and brightness contrast.** Though we mainly focus on the usage of colors, ignoring its spatial arrangement over the first two subsections, spatial correlation of colors is also important to understand the artistic style of the paintings, as shown in previous RG analysis, because a painting is a composition of colors in the proper place. The spatial arrangement of colors makes various artistic effects possible. For example, contrast, as one of the artistic effects, is an important element to express shape and space in two dimensional fine arts. Among various types of contrast, brightness contrast is the most important in art history due to the cultural background of Europe which usually adopts the contrast of light and darkness as a metaphorical expression. In this subsection, taking both the color information of pixels and their spatial arrangement into account, we examine the prevalence of brightness contrast in European paintings over 10 artistic periods.

To quantify brightness contrast, we utilize the two-point height difference correlation (HDC) and its roughness exponent  $\alpha$ , the slope of HDC curve in a double logarithmic plot of the surface growth model in statistical physics<sup>22</sup>. First we get the brightness in grey-scale from the RGB color information through a weighted transformation (see Methods), and define a "brightness surface" of an image by adopting the brightness of a pixel as a height at that position of the image as shown in Fig. 4a and b. A three-dimensional surface, like a deep-pile carpet, is obtained from the 2-dimensional painting, where the HDC is calculated as a function of distance  $r$ . This method is widely used in condensed matter and statistical physics to analyze the roughness of a growing surface, for example a semiconductor surface grown by chemical deposition<sup>22</sup>. For comparison, a shuffled image, by changing a pixel's position randomly, is analyzed together.

As shown in Fig. 4a and b, since the brightness of a point is defined as its height, the height difference between two points represents the brightness difference. The two-point HDC of a randomly shuffled painting is displayed in blue dots in Fig. 4c and d for comparison. The slope  $\alpha$  for randomized images is 0 since there is no spatial correlation any more. Figure 4d shows an example of Jackson Pollock's drip painting, which is hard to distinguish from randomly shuffled painting when only the spatial correlation is considered. The roughness



**Figure 3 | Spatial renormalization of original and shuffled images.** (a) An example of transforming an image into a fixed point. (Figure 1a also contains the image which is out of copyright.) (b) An illustrative example of the center of mass (CM), the fixed point (FP), and the shuffled fixed point (SFP) in RGB color space. (c) Norm of difference of  $d_1$  and  $d_2$  over 10 periods and comparison with Pollock's drip paintings and Pieter Bruegel the Elder's paintings. (d) Norm of cross product of  $d_1$  and  $d_2$  over 10 periods and comparison with Pollock's drip paintings and Pieter Bruegel the Elder's paintings.

exponent of Jackson Pollock's drip painting is very small comparing to that of other European paintings.

Since HDC describes the spatial correlation between color pixels on a surface as a function of distance, the slope of the HDC function, i.e., the roughness exponent  $\alpha$ , denotes the average brightness difference according to the contrast effect. Figure 5a shows that the roughness exponent  $\alpha$  gradually increases over the 10 artistic periods, which is consistent with historical circumstances. First, the increasing tendency of  $\alpha$  is related to changes in painting techniques and genres, such as from portraits to landscape. In the history of western art, many new painting techniques were developed and spread during the Renaissance period. For example, *chiaroscuro*, which is one of the canonical painting modes in the Renaissance period<sup>16</sup>, characterizes strong contrasts between light and shade. The roughness exponent and the HDC capture the level of brightness and relative spatial position. Hence, a roughness exponent  $\alpha$  of a painting could be a quantitative indicator of a *chiaroscuro* technique, and its increasing tendency over artistic periods reflects the spread of the *chiaroscuro* technique over the continent<sup>21</sup>. In addition, the Renaissance art movement led that painting genres became more diverse. Therefore, more portraits and landscape paintings were encouraged. Large objects in paintings such as a torso, i.e., the upper body of portraits, or mountains and sky in landscapes decrease the brightness difference in a short distance, but makes the increment of the HDC bigger as distance increases<sup>21</sup>. Therefore, the historical renovation of painting techniques and the diversification of painting genres are clearly captured in an increasing tendency of the roughness exponent  $\alpha$ .

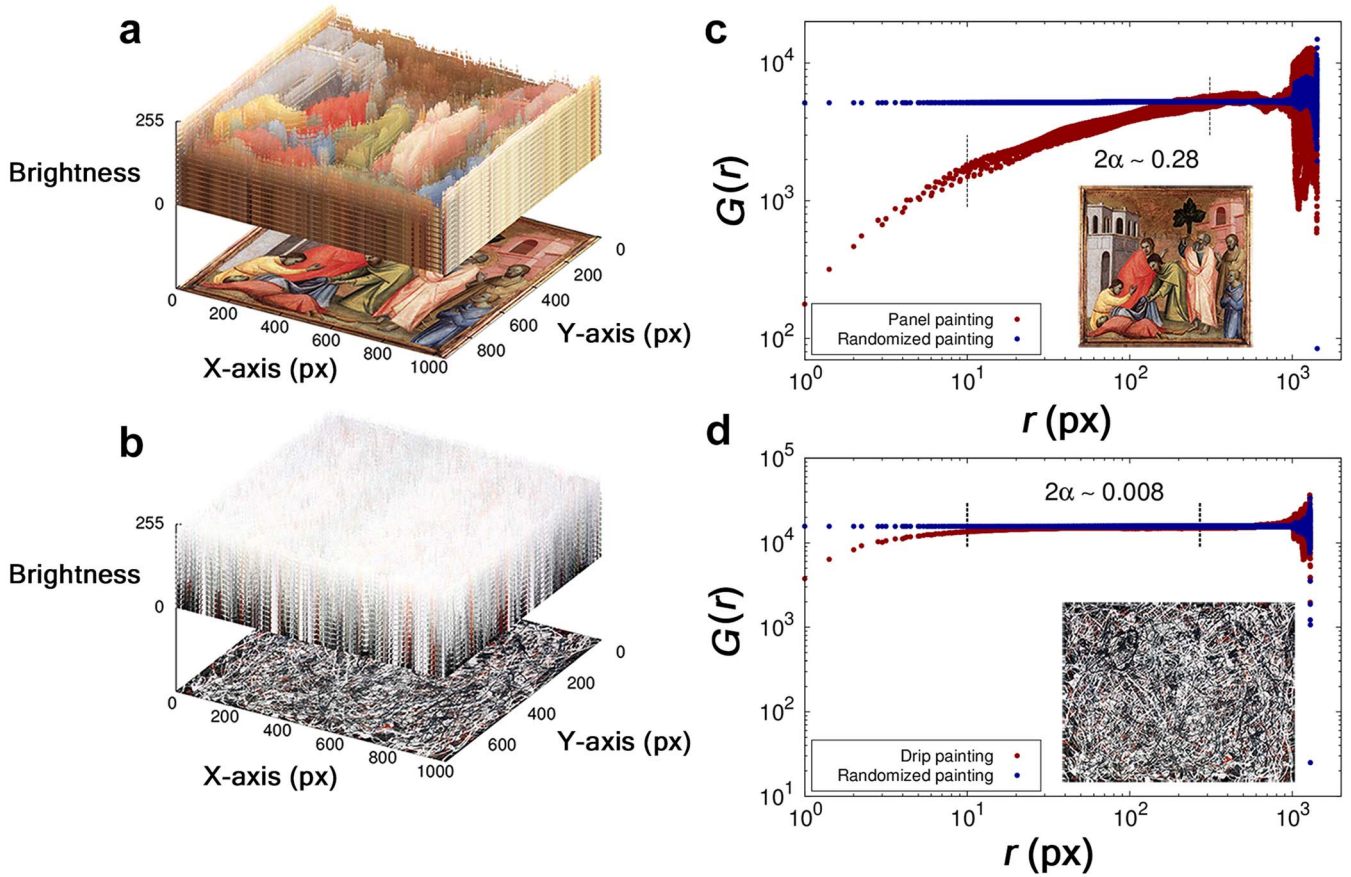
Another example, *sfumato* is another major painting mode developed in the Renaissance period to express a vanishing or shading around objects in a painting<sup>17</sup>. Smoothing the edges of objects in a

painting makes the variance of brightness decrease because it doesn't allow abrupt changes at the boundary. In this case, image entropy<sup>23</sup> would be a good measurement for the *sfumato* technique, which indicates the variance of brightness in a specific locale. Since the variance is inversely proportional to homogeneity, the image entropy describes the level of local homogeneity of brightness in a painting.

Figure 5b shows that the image entropy  $H$  increases up to Neoclassicism and then decreases, which is somewhat different from the roughness exponent since the image entropy only considers the complexity of the color gradient around a pixel locally comparing to the fact that the roughness exponent also consider the color brightness difference of remote distance. We think that the different behaviors of these two measures may reflect the tendency that the *chiaroscuro* technique is still developing but the *sfumato* declines. It may be rejecting mysterious expression and respecting the realistic one.

## Discussion

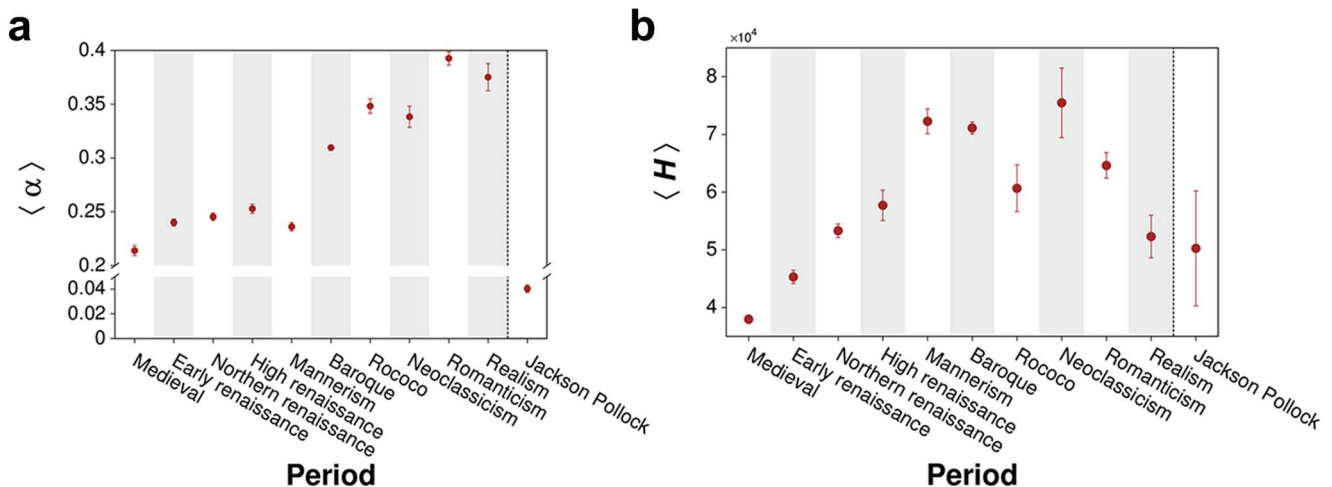
From the analysis of a large-scale European painting image archive, we display that chromo-spectroscopy of 10 art historical periods shows a universal distribution curve which distinguishes art paintings from photographs. Additionally, fractal analysis allows us to rediscover the expansion of the color palette after the medieval period, which is consistent with the fact that the color palette of the medieval age was relatively narrow comparing to other periods because of historical circumstances. Furthermore, we measure the roughness exponent and image entropy of brightness surfaces over the 10 art historical periods. We find that these mathematical measurements quantitatively describe the birth of new painting techniques and their increasing use. Our approaches successfully



**Figure 4 | Constructing brightness surfaces and measuring roughness exponents.** (a) and (b) Illustrative examples of brightness surfaces. The brightness of each point is considered as its height. (c) An example of a two-point HDC function  $G(r)$  on the brightness surface of an image in the inset, a panel painting of Italian painter Taddeo Gaddi (1348–1353) titled “*St John the Evangelist Drinking from the Poisoned Cup*” (This image is out of copyright.). The horizontal axis indicates the distance  $r$ , where a unit is a pixel, between two distinct points on the surface. Red points show the HDC of an original image and blue ones represent that of a randomized image. The slope is approximately  $2\alpha \sim 0.28$ . (d) The HDC function for an image shown in the inset, painting of American painter Jackson Pollock (1912–1956) titled “*Number 20, 1948, 1948*” (This image is reproduced by permission of the Artists Rights Society and Society of Artist’s Copyright of Korea, © 2014 The Pollock-Krasner Foundation/ARS, NY - SACK, Seoul), showing no difference from a randomly shuffled image only except for short distance less than 10 pixels, which is less than 1% of the image width.

provide quantitative indicators reflecting historical developments of artistic styles. Applying them, it is possible to deduce that the Jackson Pollock’s drip paintings are not typical art work, of course, these are still controversial in the art world.

There are several limitations of our approaches and we provide suggestions for future works. First, although the database is quite large, our dataset does not cover all paintings of the 10 art historical periods. In this reason, it is possible that there exist sampling bias in



**Figure 5 | The trend of roughness exponents and image entropies.** (a) The trend of roughness exponents over 10 art historical periods shows increasing behavior. (b) Statistical tendency of image entropy values of brightness surfaces over the periods; error bars indicate the standard error of the mean.



our results which we have not yet figured out. For better statistics, analyzing much bigger (higher resolution) images such as the Google Art Project<sup>24</sup> will give us more concrete insight for artistic style. Another possible error is unintended color distortion while converting original paintings into digital images, which may cause color information loss or bias. Even though we have checked that our results are not significantly changed from artificial color quality reductions, we could not follow all possible distortion effects. It is also true that present colors in the paintings are different from the original ones when they were completed. Old paintings are hard to preserve and usually suffer from degradation of physical materials of paintings such as oxidation and corrosion. These are big remaining issues not only for this study but also for all stylistometric analyses in arts. Nonetheless, we expect that our quantitative study would be helpful to bridge the gap between art and science.

## Methods

**Source of dataset and statistics of paintings.** In this study, we analyzed the digital images of European paintings in the Web Gallery of Art which exhibits artworks ranging from 11th century to mid-19th century<sup>15</sup>. The European paintings are classified into 10 art historical periods: medieval, early renaissance, northern renaissance, high renaissance, mannerism, baroque, rococo, neoclassicism, romanticism, and realism. We filtered non-painting images, such as sculptures, miniatures, illustrations, architecture, pottery, glass paintings, and wares. The number of refined images for each period is summarized in SI Table S1. In total we have analyzed 8,798 painting artworks. As shown in Fig. S1, over 94% of images are larger than  $700 \times 700$  pixels and the largest one is  $1350 \times 1533$ . Therefore, the quality of the images is good enough to perform a statistical analysis. Furthermore, in order to discuss the difference between paintings and photographs, two more datasets are collected for hyper-realism and photographs. We collected 105 hyper-realism images from hyper-realism artists' web sites<sup>25–31</sup>, the largest one is  $2974 \times 1954$ , and the two sets of photographs from the official Instagram site of National Geographic<sup>32</sup> and the online photo gallery of a Korean portal site<sup>33</sup>.

**Box-counting dimensions.** In order to investigate the fractal patterns of painted colors in the RGB color space, we measured box-counting dimensions<sup>20</sup>. The box-counting dimension is defined as the following:

$$d_{\text{box}}(\varepsilon) \equiv -\lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log(\varepsilon)}, \quad (1)$$

where  $N(\varepsilon)$  is the number of non-empty boxes and the side length of each box is  $\varepsilon$ . A  $\varepsilon$  value represents the color quality in a digitized unit, for example,  $\varepsilon=1$  corresponds to  $256^3$  possible colors in 24-bit RGB color system and  $\varepsilon=32$  is associated with  $8^3$  possible colors in 8-bit RGB color system. Each  $\varepsilon$  value corresponds to  $\log_2(256/\varepsilon)^3$ -bit RGB color system. Changing  $\varepsilon=32, 16, 8, 4, 2$ , and  $1$  (see Figure S6 in the supplement) and examining  $N(\varepsilon)$  for each  $\varepsilon$ , we measured  $d_{\text{box}}(\varepsilon)$ .

**Gray-scale transformation.** To consider brightness surfaces of images, we converted digital color images into grayscale images using the following weighted filter:

$$I_{\text{gray-scale}} = 0.3008 \times R + 0.5898 \times G + 0.1094 \times B, \quad (2)$$

where  $R, G$ , and  $B$  are the red, green, and blue intensities of a pixel, and  $I_{\text{gray-scale}}$  is the brightness of a certain color, which is interpreted as a height on the image. The reason for the difference in weighting values is due to the color sensitivity of a human eye<sup>34</sup>, and there exist several other weighting filters for  $R, G$ , and  $B$  intensities for specific purposes. However, there was no significant difference in the results with different filters.

**Two-point height difference correlation function.** To measure the roughness exponents of brightness (height) surfaces, a two-point height difference correlation (HDC) function is calculated<sup>22</sup>. The definition is

$$G(r) = \overline{[h(\vec{x} + \vec{r}) - h(\vec{x})]^2} = \frac{1}{N_r} \sum_{\vec{x}, |\vec{r}|=r} [h(\vec{x} + \vec{r}) - h(\vec{x})]^2, \quad (3)$$

which follows the simple scaling form,  $G(r) \sim r^{2\alpha}$ , for small  $r$ , and where  $r$  is a distance between two pixel points, the over-bar represents the spatial average at a fixed distance  $r$  for all possible points,  $N_r$  is the number of possible pairs at a distance  $r$ ,  $h(x)$  is the height at a point  $x$  ( $0 \leq h(x) \leq 255$ ), and  $\alpha$  is the roughness exponent. The roughness exponent was measured in a double-logarithmic plot of  $G$  versus  $r$ , where the fitting range was used from  $r_a=10$  to  $r_b$ , where the HDC saturates to the same value both for the original and randomized paintings. It approximately corresponds to 30% of the image width and a square root of 9% of the image area.

**Image entropy.** Entropy of a gray-scale image<sup>23</sup>, is given by the following equation:

$$H = -\sum_x p(x) \log \frac{p(x)}{m(x)}, \quad (4)$$

where  $p(x)=h(x)/S$ ,  $h(x)$  is the height at a point of the brightness surface ( $0 \leq h(x) \leq 255$ ) and  $S$  is the sum of all height values in the image for normalization. A weighting factor  $m(x)$  is given by  $m(x)=1+\sigma^2(x)$ , where the local height variance

$\sigma^2(x) = \frac{1}{9} \sum_y \left[ h(y) - \frac{1}{9} \sum_y h(y) \right]^2$  is calculated only over for its surrounding

neighbor pixels and itself at a position  $x$ . Since this image entropy depends on an image size, all images are resized to  $500 \times 500$  pixels by Lanczos algorithm before measuring the image entropy.

- Taylor, R., Micolich, A. & Jonas, D. Fractal analysis of Pollock's drip paintings. *Nature* **399**, 422 (1999).
- Lyu, S., Rockmore, D. N. & Farid, H. A digital technique for art authentication. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17006–17010 (2004).
- Johnson, C. R. Jr. *et al* 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**, 37–48 (2008).
- Bressan, M., Cifarelli, C. & Perronin, F. An Analysis of the relationship between painters based on their work. *15th IEEE Int. Conf. Image Proc.* 113–116 (2008).
- Olshausen, B. A. & DeWeese, M. R. Applied mathematics: The statistics of style. *Nature* **463**, 1027–1028 (2010).
- Hughes, J. M., Graham, D. J. & Rockmore, D. N. Quantification of artistic style through sparse coding analysis in the drawings of Pieter Bruegel the Elder. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1279–1283 (2010).
- De Morgan, S. E. *Memoir of Augustus de Morgan, by his wife Sophia Elisabeth de Morgan, with Selection of his Letters* (Longmans, London, 1882).
- Lutostowski, W. *The Origin and Growth of Platons Logic* (Longmans, Green, London, 1897).
- Holmes, D. I. & Kardos, J. Who was the author? An introduction to stylometry. *Chance* **16**, 5–8 (2003).
- Hughes, J. M., Foti, N. J., Krakauer, D. C. & Rockmore, D. N. Quantitative patterns of stylistic influence in the evolution of literature. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 7862–7866 (2012).
- Manaris, B. *et al*. Zipf's law, music classification, and aesthetics. *Comput. Music J.* **29**, 55–69 (2005).
- Huron, D. The ramp archetype: A study of musical dynamics in 14 piano composers. *Psychol. Music* **19**, 33–45 (1991).
- Casey, M., Rhodes, C. & Slaney, M. Analysis of minimum distances in high-dimensional music spaces. *IEEE Trans. Speech Audio Proc.* **16**, 1015–1028 (2008).
- Sapp, C. Hybrid numeric/rank similarity metrics for musical performances. *Proc. ISMIR* **99**, 501–506 (2008).
- Krén, E. & Marx, D. Web Gallery of Art, image collection, virtual museum, searchable database of European fine arts (1000-1900) <http://www.wga.hu/> (1996) Date of access 09/05/2009
- The National Gallery, London: Western European painting 1250–1900 <http://www.nationalgallery.org.uk/paintings/glossary/chiaroscuro> Date of access 03/04/2013.
- Earls, I. *Renaissance Art: A Topical Dictionary* (Greenwood Press, 1987).
- Zipf, G. K. *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology* (Addison-Wesley Press, Cambridge, 1949).
- Abramowitz, M. & Stegun, I. A. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables* (Dover Publications, New York, 1965).
- Gouyet, J. F. & Mandelbrot, B. *Physics and Fractal Structures*. (Springer-Verlag, New York, 1996).
- Gage, J. *Color and Meaning: Art, Science, and Symbolism* (University of California Press, Berkeley, 2000).
- Barabási, A. L. & Stanley, H. E. *Fractal Concepts in Surface Growth* (Cambridge University Press, Cambridge, 1995).
- Brink, A. D. Using spatial information as an aid to maximum entropy image threshold selection. *Pattern Recogn. Lett.* **17**, 29–36 (1996).
- Google Inc. Google Cultural Institute <http://www.google.com/culturalinstitute/project/art-project> (2011) Date of Access 14/02/2011
- Jacques Bodin gallery <http://www.jacquesbodin.com/> Date of Access 07/04/2010.
- Roberto Bernardi <http://www.robortobernardi.com/> Date of Access 07/04/2010.
- Raphaella Spence <http://www.raphaellaspence.com/> Date of Access 07/04/2010.
- Hubert de Lartigue Accueil <http://www.hubertdelartigue.com/> Date of Access 07/04/2010.
- Gus Heinze on artnet <http://www.artnet.com/artist/26318/gus-heinze.html> Date of Access 07/04/2010.
- Bernardo Torrens <http://www.bernardotorrens.com/> Date of Access 07/04/2010.
- The Hyperrealism Paintings by Denis Peterson <http://www.denispeterson.com/> Date of Access 07/04/2010.
- National Geographic official Instagram. <http://instagram.com/natgeo> (1999) Date of access 20/08/2014



33. Naver, officially launched in 1999, was the first Korean portal site to develop its own search engine. <http://www.naver.com/> (1999) Date of access 20/07/2009
34. Pratt, W. K. *Digital Image Processing* (John Wiley & Sons, New York, 1991).

## Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Ministry of Science, ICT & Future Planning (No. 2011-0028908).

## Author contributions

D.K. designed and performed research, analyzed data and wrote the paper; S.-W.S. designed and performed research, and wrote the paper; H.J. designed research and wrote paper. All authors discussed the results and commented on the manuscript.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Kim, D., Son, S.-W. & Jeong, H. Large-Scale Quantitative Analysis of Painting Arts. *Sci. Rep.* 4, 7370; DOI:10.1038/srep07370 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>