Abstract

This paper suggests that tools used for the study of the statistics of natural images can be fruitful in the context of computer depiction. Such techniques study the regularities of natural signals and capture relevant features to characterize or classify various ensembles of images. They have been shown to be related to the human visual system. We propose that picture statistics can be used to characterize simple notions of pictorial style, and that in the future, they might lead to classification, style capture, or picture manipulation and enhancement tools.

After reviewing the use of image statistics in vision science and style classification for images and music, we study the frequency and multiscale properties of various pictorial styles. We study the average amplitude of coefficients of multiscale decomposition, and show that their variation with respect to scale correlates with a simple aspect of pictorial style.

Keywords: image processing, statistics of natural images, steerable pyramids

1 Introduction

Pictorial style is an elusive notion that can be approached from many different angles, such as historical, iconographic or pictorial. Most work on pictorial style has relied on (subjective) written analysis, but rarely on objective quantitative measurements. We believe that quantitative notions of style are important for computer depiction, and we suggest that the field of the statistics of natural images has developed a set of tools that can lead to such objective style criteria. We must emphasize that we do not claim that they will capture all aspects of style, and that they will certainly not capture the genius or the notion of art. Nevertheless, they will afford a better understanding of some dimensions of style as well as important practical tools for computer depiction.

The intuition behind our work is that some pictorial styles differ in very dramatic ways and that they can be distinguished even at a pre-attentive level, that is, without carefully scrutinizing the picture. For example, the difference between a Pointillist painting and a Renaissance picture is blindingly obvious because the former contains much more fine-grain texture due to the juxtaposition of contrasted dots. Similarly, a photograph by Ansel Adams differs at the first sight from the gray-scale version of a casual photograph because of the strong tonal contrast between regions of the image. There are of course other more subtle dimensions to style, but these strong differences are a first step to study this complex notion.

In order to study the regularities within a style and the differences between styles, we propose to use statistical tools developed for the study of natural images. It has long been known that not all random two-dimensional signals look like natural images. In recent years, a wealth of research has addressed the characterization of natural signals. Various statistical tools have been used to capture the regularities encountered in natural signals. It is hypothesized that these statistics might relate to the characteristics of the human visual system either because vision has evolved to leverage these regularities, or because these statistics are simple to gather and can facilitate early recognition and classification of various phenomena. We give an overview of this work in Section 1.1.

We believe that objective and quantitative tools that capture notions of pictorial style can have several applications in photorealistic and non-photorealistic computer graphics. They can be embedded in image classification and retrieval tools. They can facilitate the separation of style and content, allowing for direct style transfer, or for the capture of style in the context of NPR systems driven by complex parameters. For example, the parameters of a painterly 3D renderer could be inferred to match the statistics of a given input image. In the context of a system like design galleries [Marks et al. 1997] where the user explores a complex parameter space, statistical criteria could be used to maximize the diversity of styles proposed to the user. For digital photography, they could be used for style transfer to obtain some of the basic qualities of master photographers, or to drive image enhancement tools such as tone mapping. In addition, we hope that the statistical dimensions revealed by such studies might lead to novel direct image manipulation tools that act directly on basic stylistic aspects. For example, we will see that our initial study suggests that an interactive control over the contrast at different scales leads to interesting image manipulation interface. Finally, statistical estimators might in the future help understand what makes images photorealistic, and lead to simple filtering tools to increase realism [Rademacher et al. 2001; Leykin and Cutzu 2002; Cutzu and Hammoud 2003]. We are far from these ambitious goals, but we believe that our initial study lays out the path and shows promising results.

1.1 Related work

Statistical tools have received little attention in computer depiction. As discussed above, we believe that they can stimulate fruitful research by providing measurable notions of style. We first briefly review the treatment of style in non-photorealistic rendering, before introduce the statistics of natural images and other statistical and perceptual studies of style in images and music. The study of these related work provides much inspiration for our field, but also should lower our expectation for the granularity of style characterization. Results from related fields such as musical style suggest that initial result will be limited to coarse-grain notions of style.

Style in computer depiction

Most Non-photorealistic rendering approaches model style as an integral aspect of the algorithm, e.g. [Winkenbach and Salesin 1994; Salisbury et al. July 1994; Ostrovomkhov 1999; Kowalski et al. 1999; Hertzmann and Zorin 2000], as a set of parameters [Hertzmann 1998], or as a result from user interaction [Haebi et al. 1990; Schofield 1996; Kahn et al. 2002].

Recently, approaches inspired from markov fields and information theory [ ] have permitted the transfer of the textural aspect of style, by mimicking the local texture between a source and destination image [Hertzmann et al. 2001; Efros and Freeman 2001; Drori et al. 2003] or set of strokes [Hertzmann et al. 2002; Jodoin et al. 2003].
Statistics of natural images

We give a brief overview of the studies on the statistics of natural images. We refer the interested reader to the more comprehensive survey by Simoncelli and Olshausen [2001] or to the technical report by Reinhard et al. [2001b].

The roots of the statistical study of image scan be traced to the hypothesis by Attnave [1954] and Barlow [1959] who build on information theory to hypothesize that the human visual system takes advantage of regularities in natural stimuli to compress this information and extract more compact representations. This can be linked to the ecological optics theories of Gibson who suggests that the study of vision must be grounded on the available cues present in the visual field [Gibson 1950; Gibson 1979].

Statistical properties of images are usually studied by collecting a number of such images in an ensemble and computing first, second or higher order statistics. First-order statistics treat each pixel independently. Second-order statistics measure the dependencies between pairs of pixels. Higher order statistics are used to extract properties of natural scenes that cannot be modeled using first and second-order statistics. Statistics can be computed on teh image intensities themselves, or on Fourier transforms, wavelet coefficients, steerable pyramids, or the response to preceptually-inspired filter banks such as Gabor functions.

The first statistical studies of natural images have focused on the Fourier spectrum. Many natural processes exhibit power spectra with 1/f characteristics. For natural images, the spatial power spectra from natural images falls approximately as the square of the spatial frequency and this property is called “scale invariance” [Ruderman and Bialek 1994]. This means that a natural scene has equal amounts of structure at all scales. In Fourier terms, this implies that any spatial-frequency band with the same octave bandwidth will have an equal amount of energy. It was also noted that contour plots of the power spectrum of the natural-scene log contrast have shown to share a preponderance of power in low spatial frequencies along the horizontal and vertical orientations [Ruderman 1994; Torralba and Oliva 2003], because of preponderant features such as the horizon line or vertical tree trunks and the axis-aligned edges of man-made objects. Unfortunately, Fourier analysis is not optimal for non-stationary signals, and edges are not localized in Fourier space.

Since the Fourier spectrum of natural images are not stationary, the wavelet transform has provided a powerful alternative. The wavelet transform has the good property that the coefficients of wavelet subbands are nearly decorrelated [Bucciogrossi and Simoncelli 1997b] and the magnitudes of pairs of wavelet coefficients are highly correlated. The main result on the natural statistics of wavelet coefficients is that their distribution is non-Gaussian and always show a typical long-tail distribution (Fig. 1).

Statistical moments are used to characterize such a distribution. The first (mean) and second (variance) moments intuitively characterize the x and y scaling of the plot. The following moments are by definition zero for Gaussian distributions. The third moment (skewness) tells how symmetric around the peak the distribution is. Finally the fourth moment (kurtosis) describes how pointy the data is. It is usually defined so that the kurtosis of a Gaussian is zero (a.k.a excess kurtosis). Negative kurtosis indicates a flat distribution, while positive kurtosis indicate peaks and heavy tails. Natural images always exhibit a high kurtosis.

We will come back to these properties in relation to paintings in section 2. But the orthonormal wavelet representation suffers from a lack of translation-invariance (non-shiftable) [Simoncelli et al. 1992]. This means that the histogram of coefficients of a wavelet transform vary under translation. In addition, the discontinuous basis used in wavelets such as Haar are likely to cause artifacts in an application like image manipulation or texture synthesis.

Finally Hamel and Strothotte [1999] propose to learn the parameters defining a style for a non-photorealistic 3D renderer from an example specified by the user. They use templates that capture the mapping between features of the 3D models and the output picture. We believe that image statistical tools can in the future greatly facilitate this process.

Figure 1: The marginal distribution of wavelet coefficient for natural images from http://www.cis.hut.fi/projects/ica/data/images/

The steerable pyramid [Freeman and Adelson 1991; Simoncelli et al. 1992; Simoncelli and Freeman 1995] is a good substitute for the wavelet transform. This transform has nice reconstruction properties (specifically, it is a tight frame), in addition to properties of translation-invariance and rotation-invariance. Steerable transform shares the same conceptual structures with the wavelet transform. Its statistical properties usually follow those of the wavelet transform. We recommend the use of steerable pyramids for graphics applications because they offer better reconstruction.

Statistics of other visual stimuli such as natural range images [Huang et al. 2000] and natural illumination [Dror et al. 2001b] have also been studied, and similar regularities were found.

Link to vision [Field 1987; Simoncelli and Olshausen 2001]

color statistics

Most of the statistics described so far were computed on gray-scale images. The statistics of color signals were studied both in the full wavelength spectrum [Maloney 1986] and after projection onto the...
Applications of image statistics

The statistics of natural image are important tools because they allow one to restrict the set of possible output images for a given operation. They provide a prior knowledge on the set of possible outputs. For example, in denoising, enforcing known regularities in the statistics of natural images results in outputs that are more likely to be accurate [Portilla et al. n. d.; Simoncelli 1997]. Similarly, for image compression, regularities can be exploited for better prediction and therefore better compression, e.g. [Buccigrossi and Simoncelli 1997a; Simoncelli 1997]. Superresolution can also greatly benefit from priors on the magnified image [Sun et al. 2003].

Previous application rely on statistical properties that are shared by all images. In contrast, statistical texture synthesis applications attempt to characterize the precise statistics of an individual texture. [Lewis 1984; Heeger and Bergen 1995a; Zhu and Mumford 1997; Portilla and Simoncelli 2000a]. Similar textures are then synthesized by enforcing the statistics on an input white noise. Such texture models can also be used for inpainting image restoration, e.g. [Hirani and Totsuka 1996; Igehy and Pereira 1997]. In his PhD thesis, page 125, Debevec [1996] mentions that he considered using such techniques on Monet’s Rouen series for hole filling. Note that statistical texture synthesis is different from recent non-parametric models, e.g. [Efros and Leung 1999; Efros and Freeman 2001], in that they also suggest a similarity metric between textures. The validation of such metrics is however still an area of active research.

Image statistics provide a powerful means to learn priors from example, leading to numerous computer vision applications, e.g. [Freeman et al. 2000; Tappen et al. 2003; Dror et al. 2001a; Moghaddam et al. 2003], as well as image classification, e.g. [Torralba and Oliva 2003]. We will discuss work in the domain of painting classification shortly.

Reinhard et al. [2001a; 2001b] introduced the statistics of natural images into computer graphics. They focused on the 1/f characteristic of the Fourier spectrum and studied which aspects of digital image synthesis might affect the slope of the spectrum, and therefore the naturalness of images [Reinhard et al. 2001b]. They concluded that modeling is the key step affecting the frequency spectrum. They also used the results by Ruderman [1999] on the statistics of colors to derive a color matching technique [Reinhard et al. 2001a]. In this article, we introduce a new set of statistical tool and study not only natural images, but also paintings.

Computer depiction and the separation of style and content lie somehow between the need for generality encountered in the study of natural images, and the precise characterization of particular categories found in texture analysis/synthesis. In many instances, we want the content to be treated as a regularity similar to that of natural images, but to capture the precise characteristics of a given style, whatever we mean by style.

Statistical study of style for other modalities

Statistical tools have also been successfully used to characterize style in text, e.g. [Sebastiani 2002] and in music. We focus on the latter because the continuous nature of the signal makes it more similar to images. We believe that the computer depiction community can find much inspiration in the work done on music statistics.

Musical style has been studied at various granularities: the score (discrete), the performance, the timbre of individual sounds, and the style of an entire input digitized sound piece. We do not discuss the style of the score itself, because it is a discrete issue not really related to images. Although it is also essentially a discrete problem, we highly recommend the reading of a paper on the link between score and performance. Indeed, Widmer et al. [n. d.] presents a very thorough discussion of the possible goals and evaluation of research related to style and artistic qualities.

The most famous result on the study of musical signals is that they exhibit a 1/f spectrum [Voss and Clarke 1975; Voss and Clarke 1978], in contrast to Brownian-motion signals or noise. Recently, statistical classifier of musical style are developed, mostly for music file retrieval applications to help users find songs similar to the ones they like, e.g. [Dannenberg et al. 1997; Welsh et al. 1999; Logan and Salomon 2001; Tzanetakis and Cook 2002]. They can also be used for computer accompaniment, where the computer reacts to the style of the human musician [Dannenberg et al. 1997; Dannenberg 2000]. Most approaches use spectral features and learn from labeled data. They usually try to classify between four to a dozen different style, which gives a good idea of the granularity of style (e.g. jazz vs. classical vs. rock). The success rates are impressive but far from perfect, which should lower our expectation to find soon perfect pictorial style classifier able to recognize individual artists.

In contrast to the study of an entire piece, some techniques characterize the “timbre” of individual instruments. [Wessel 1979; Dubnov et al. 1995; Dubnov 1996]. They use similar spectral features, but the characterization is finer because sounds within one category (one instrument) are quite similar. For example, Dubnov and Tishby [1997] use higher-order statistical properties of the spectrum. They show that the skewness (3rd moment) and kurtosis (4th moment) locate the instruments according to traditional music handbook practices. The decoupling of timbre from other sound aspects such as pitch, loudness and brightness can then lead to powerful synthesis interfaces, e.g. [Jehan and Schoner 2001; Wright et al. 1999].
Other style studies

As discussed above, the statistics of natural images relate to properties of the human visual system. Similarly, many authors have related pictorial style to vision sciences, e.g. [von Helmholtz 1881; Gombrich 1956; Arnheim 1954; Solso 1994; Ramachandran and Hirstein 1999; Zeki 2000; Livingstone 2002; Durand 2002b]. In particular Arnheim [1971] describes how style relates to order and entropy. His plate 6 and 7 actually represents the evolution of a Poussin painting under low pass filtering, which directly relates to the spectral characteristics of the painting. Solso [1994] summarizes research showing that the statistics of gaze fixation correlate directly with a notion of style complexity.

Willats [Willats 1997; Willats and Durand 2003] and Durand [Durand 2002a] describe a coarse-grain classification of style that is based on perceptual notions, but that addresses the problem from the complete opposite standpoint of teh present paper. They focus on the high-level mapping between the represented scene and the picture, while we focus on low-level properties of the 2D pictorial signal.

1.2 Approach overview

When it comes to studying a notion as elusive and complex as style, methodology is a crucial issue. We need to decide which statistical tools to use and how to evaluate success or at least relevance and correlation with style. Our work is grounded on foundations from the field of the statistics of natural images. Our methodology is inspired by work on texture analysis and synthesis, e.g. [Heeger and Bergen 1995b; Portilla and Simoncelli 2000b]. It relies on five steps:

- We use image decomposition tools that were shown to be relevant or plausible for the study of visual stimuli, such as Fourier decomposition, wavelets and steerable pyramids.
- We compute a variety of statistical estimators on these decompositions.
- We compare these properties to the regularities observed in natural images.
- We look for properties that are stable within a style but that vary across styles.
- We test if these variations capture a notion of style by transferring these properties to input images of a different style.

We still need to discuss what we mean by style. Most related approaches rely on supervised learning where input images are labelled with a given style and learning techniques are used to derive classifiers. For paintings, we decided to use a different approach where categories are not necessarily known a priori, but can be discovered by human observation of the statistical properties. We then test the hypothesis using style transfer.

We believe that the two approaches are complementary. We test the hypothesis using style transfer.

2 Statistics of paintings

2.1 Fourier spectrum

We computed the Fourier transform of these paintings and analyzed their power spectrum (Figure 2, Figure 3 and Figure 4). Figure 2 shows that the regularities found in natural images are also valid for paintings, and notably that the power spectra fall off as \(1/f^2\) [Field 1987; Ruderman 1994].

Figure 3 shows the Fourier spectrum of paintings from two artists. The red region near the center of the spectrum depicts that the main concentration of the painting’s energy is at low frequencies, which is typical in real-world images. Figure 4 points out the horizontal and vertical line of spectral energy are preferred orientations, due to salient lines such as the horizon or vertical tree trunks [Ruderman 1994; Torralba and Oliva 2003].

The differences between Monet’s spectra and Manet’s in Fig. 3 imply the Fourier spectrum reveals simple characteristics of painting style such as the size of brush touches. Manet’s paintings are smooth overall. Therefore, most of the signal energy is contained in components at the center. In contrast, the signal energy in the Monet’s paintings is placed in a relatively wider spectral band, since Monet’s paintings contain more salient high-frequency marks (brush strokes).

The Fourier transform is a useful tool for simple analysis of pictures. But like for real-world images, the statistics of paintings are not stationary over the images, and images contain many edges that are poorly characterized through Fourier analysis. We extend our analysis from Fourier transform to Wavelet transform.
The Fourier transform of Manet’s *Emile* and *Olympia* and of Monet’s *Hay* and *Poplars*. These demonstrate the paintings exhibit $1/f^2$ power spectrum (or $1/f$ amplitude spectrum).

2.2 Haar wavelets

The wavelet transform has been successfully applied to non-stationary signals and provides better locality of edges. The use of wavelet analysis to characterize natural signal has been quite successful [Buccigrossi and Simoncelli 1997a; Huang and Mumford 1999].

We have chosen The Haar wavelet basis for its simplicity. Building upon work on the statistics of Haar wavelets for natural images, we computed the same statistical properties for paintings. Similar to the Fourier analysis, we found that paintings exhibit the same overall characteristics as natural images. Figure 5 and Table 2 illustrate that the wavelet coefficients of paintings have non-Gaussian distributions, like those of natural images in [Buccigrossi and Simoncelli 1997a]. Figure 1 and Table 3 represent the typical non-Gaussian distributions of the natural images.

<table>
<thead>
<tr>
<th>Artist</th>
<th>Drawing</th>
<th>Kurt (Ver)</th>
<th>Kurt (Dia)</th>
<th>Kurt (Hor)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degas</td>
<td>Beside</td>
<td>10.2565</td>
<td>10.7184</td>
<td>10.5368</td>
</tr>
<tr>
<td>Degas</td>
<td>Cotton</td>
<td>23.5992</td>
<td>52.5444</td>
<td>39.6472</td>
</tr>
<tr>
<td>Cézanne</td>
<td>Abduct</td>
<td>17.9390</td>
<td>27.2391</td>
<td>31.8267</td>
</tr>
<tr>
<td>Cézanne</td>
<td>Approch</td>
<td>13.8478</td>
<td>9.0323</td>
<td>7.9583</td>
</tr>
<tr>
<td>Signac</td>
<td>Clichy</td>
<td>12.8637</td>
<td>16.0005</td>
<td>13.0265</td>
</tr>
<tr>
<td>Signac</td>
<td>Pine</td>
<td>10.6099</td>
<td>10.0795</td>
<td>10.3674</td>
</tr>
<tr>
<td>Bellini</td>
<td>Emo</td>
<td>9.7949</td>
<td>10.3521</td>
<td>10.8427</td>
</tr>
<tr>
<td>Bellini</td>
<td>Feast</td>
<td>19.0060</td>
<td>6.7693</td>
<td>11.1218</td>
</tr>
</tbody>
</table>

Table 2: The Kurtosis in the randomly chosen paintings from several artists

Although the paintings share the regularity of natural images, we observed that the distribution of wavelet coefficients correlates with a simple notion of style. Figure 5 displays the marginal distribution of wavelet coefficient of two paintings in the log domain.

They demonstrate the low-level difference between Manet’s painting style and Monet’s painting style. The coefficients in the Bar are quite focused on zero. This is because the painting mostly consist of smooth parts. In contrast, the marginal distribution for Monet’s *Hay* has less kurtosis, since the painting is composed with very delicate brush touches. As a results, the histogram of coefficients contains a larger number of big coefficients.

If we closely look at the statistics across scale, the notion of style is unveiled more noticeably. Figure 6 displays the means of the absolute values of the detail coefficients at the scales from 1 to 5 in the wavelet transform of the paintings. A basic but important characteristic of paintings is highlighted in the graph of the averages. Manet’s paintings have low values at the fine scale and high values at the coarse scale. Signac’s paintings have high values rather at the fine scales. This points out that Signac uses a many small points and that Manet emphasizes the large-scale contrast between large regions of the painting.

$$M_{jk} = \text{mean}(|\text{WaveletBand}_{jk}|)$$

$$j = \text{orientation}, k = \text{scale}$$

Compared with Figure 6, the plots of the natural images do not have the slopes in Figure 7. This implies the slope is something the artists generate purposely. This argument is approved in the artistic photographs. They have the slopes in Figure 8 which plots the means of the absolute values of the detail coefficients at the scales from 1 to 5 in the wavelet transform of the artistic photographs.

2.3 Joint wavelet coefficients

So far, we have studied marginal (first-order) statistics of wavelet coefficients. We now employ joint statistics that reveal more structure in images by capturing the correlation between nearby wavelet coefficients.

When we take two random variables $X$ and $Y$ from $S$ and $T$, we can think of $(X, Y)$ as a random variable taking values in the product set $S \times T$. To understand the relationship between $X$ and $Y$, it is important to know how the distribution of $(X, Y)$ is related to the distributions of $X, Y$. The distribution of $(X, Y)$ is called the joint distribution of $(X, Y)$, while the distributions of $X$ and $Y$ are referred to as marginal distributions. The conditional histogram shows the distribution of $(X, Y)$ with given $X = x$. 

$M_{jk} = \text{mean}(|\text{WaveletBand}_{jk}|)$

$\text{Joint} = \text{orientation}, \text{scale}$

$\text{Compared with Figure 6, the plots of the natural images do not have the slopes in Figure 7. This implies the slope is something the artists generate purposely. This argument is approved in the artistic photographs. They have the slopes in Figure 8 which plots the means of the absolute values of the detail coefficients at the scales from 1 to 5 in the wavelet transform of the artistic photographs.}$
This provides some validation that paintings have the same statistical relationships between pairs of coefficients.

Figure 10 is the conditional histogram \( H(C | P) \) for Signac’s Dining and for Manet’s Bar.

We also want to study the joint distribution of neighboring coefficients at the same scale.

Figure 11 is a result produced by the texture synthesis method using the joint statistics. It implies the joint statistics is related to the structure of the image.

Discuss that the effect of joint might be that they are interesting for style, or that they fix the problem with poor edge localization.

2.4 Steerable pyramids

Steerable pyramid [Freeman and Adelson 1991; Simoncelli et al. 1992; Simoncelli and Freeman 1995] have the same multi-resolution structure as wavelets. They are however better at capturing statistics of images because they are shift-invariant, and they contain more complete multi-orientation information. In addition, since their basis functions are smooth, steerable pyramids are more appropriate for image manipulation than Haar wavelets.

We found that steerable pyramid statistics are very similar to that of wavelet coefficients. Figure 12 displays the averages of the detail coefficients’ absolute values at multi-scales for paintings in the steerable pyramid domain. The graphs look really similar to the graphs in the wavelet transform domain. In the following section, we exploit the ability of steerable pyramids to capture the simple scale-based notion of pictorial style.

2.5 Color statistics

In Figure 13, we see that the RGB color space shows almost complete correlation between all pairs of axes. On the contrary, \( l, \alpha \) and \( \beta \) are decorrelated.

3 Visual equalizer

In the previous section, we have shown that a basic notion of style correlates with the multi-scale content of images. This suggests a new technique for image manipulation where the amplitude of different scales is modified. This idea is in fact quite similar to the equalizer offered on most audio systems. Just like the manipulation of various frequency bands provides a simple control on the ambiance of music, we believe that the manipulation of coefficients at different scales provides a powerful low-level image manipulation technique.

This approach can be seen as a simpler version of statistical texture synthesis [Heeger and Bergen 1995b; Portilla and Simoncelli 2000b] applied to non-textured images. Many authors have used
sets of multi-scale bandpass filters for texture synthesis [Heeger and Bergen 1995b; Portilla and Simoncelli 2000b]. We borrow the idea of histogram matching [Heeger and Bergen 1995b] proposed by Heeger and Bergen and apply it to simple low-level style transfer.

3.1 Interactive equalizer

Our system is based on a set of measurements on the coefficients of the steerable pyramid. We decided to use a steerable pyramid, since unlike the wavelet transform, it does not exhibit block artifacts when the coefficients are modified. In addition, it allows adaptive control over phase as well as orientation.

We chose to use the $l\alpha/\beta9$ color channel since each channel is decorrelated.

We have implemented an interactive visual equalizer where the user controls through sliders the gain applied to the various scales (see Figure TODO). As soon as it is implemented, we will love it and we will prove that it is great.

3.2 Histogram matching and style transfer

The texture synthesis algorithm developed by Heeger and Bergen [Heeger and Bergen 1995b] is based on an iterative sequence of histogram-matching operations. It was later improved by Portilla and Simoncelli [2000b] using more comprehensive statistics. We apply the idea to our simple multi-scale histogram of average steerable coefficients.

As we have shown in the previous section, the averages of the detail coefficients' absolute values at multiple scales represent a clear trend for each broad category of painting style. We use the histogram of average coefficient for each scale as a simple model for the “ambiance” of a painting or set of paintings. We can then modify an input image $B$ so that it matches the histogram of a reference image $A$ simply by multiplying the coefficients of $B$ at each scale by the ratio of the histogram values from $A$ and $B$.

$$\text{coefficient}^a_{jk} = \frac{B_{jk}}{A_{jk}}$$

$$A_{jk} = \text{mean}(|\text{spyrBand}^a_{jk}|)$$

$$B_{jk} = \text{mean}(|\text{spyrBand}^b_{jk}|)$$
Figure 8: A plot for the means of the absolute values of the detail coefficients at the scales from 1 to 5 in the wavelet transform of the artistic photograph. We can see each plot has the slope that the plots of paintings had.

Figure 9: The conditional histogram $H(\log_2(C) \mid \log_2(P))$ for horizontal coefficients. Bright parts mean high values. The pair of wavelet coefficients are correlated at the high values.

Figure 10: The conditional histogram $H(C \mid P)$ for horizontal coefficients. Bright parts mean high values.

Figure 11: The results from the texture synthesis using joint statistics after 25 iterations in the gray scale. This shows the joint statistics is related to the structure of the image.

We demonstrate this technique using two extreme reference histograms. Using the 8 paintings by Manet in our dataset, we define a Manet histogram as the average of the 8 histograms. In fact, we compute 3 histograms, one for each color channel. We similarly extract a Signac histogram. Figure 14 shows the histograms. We use these reference histograms to generate two equalized paintings in Figure 15 from an input painting by Pissaro, *Mathurins*. The middle picture was matched to the Manet histogram and the lower one to the Signac histogram. We can see that the Signac matching sharpens the details and that the Manet matching smoothes the details. This is because the value at the finest scale in the Signac histogram is larger than in the Manet histogram. Matching Manet’s histogram emphasizes the contrast between objects that compose the painting, since it has high value in the coarse scale (Fig. 16).

$$M_{j,k} = \text{mean}(|\text{spyBand}_{j,k}|)$$
Figure 12: The mean of the absolute values of the detail coefficients from the steerable pyramid in the gray scale. From top to bottom, Manet’s Bar, Manet’s Émilie, Signac’s House and Signac’s Dining. These show the same trends as they did in the wavelet transform.

\[ M_{jk} = \sum_{i=1}^{n} M_{ijk} \]

\( i = \text{painting}, n = \text{the number of paintings}, \]
\( j = \text{scale}, k = \text{orientation} \)

Matching based on the RGB color space only provides a simple extension to gray scale notions. It does not really affect the hue and saturation of colors, because tone and color are poorly decorrelated. To modify color tones effectively, we choose the Lab color space [Ruderman et al. 1999].

Figure 13 compares the results of applying histogram matching in RGB and in Lab color space. To generate the reference histogram, we take the average of each painting’s average value for RGB and the Lab color channels. This is still done at multiple scales. It is noticeable that the upper image and the middle image have the same color tone. Comparing the lower image with the middle and the upper images reveals the color tone of the modified image is different from the original painting.

Figure 14: Manet-model and Signac-model: Scale 1 means the finest scale.

4 Discussion

With Fourier transform, wavelet transform and steerable pyramid, we have shown that the statistics of paintings are consistent with those of natural images, and that they correlate with simple notions of pictorial style. We use different image transform tools to analyze the characteristics of paintings. Also, by using simple multiscale equalization, we develop a novel method of editing paintings.

Our method captures an extremely simple notion of style. It does not have the ability of image analogies [Hertzmann et al. 2001] to reproduce fine texture details. However, it better models the large-scale contrast that influence picture composition. We hope that more complete statistics such as that used in statistical texture synthesis will be able to capture more complex notions of pictorial style. We also want to use these statistics and machine-learning tools to build automatic pictorial style classifiers. We also want to study less realistic pictorial styles and in particular line drawing or hatching, since their statistical content must be dramatically different from that of paintings due to the large number of high-contrast thin strokes.

Our vision: properties get refined, more style discrimination, better and better transfer.

4.1 To do list
Study in CIE Lab color space interactive equalizer
Histogram matching in Fourier domain
Comparison between natural images and paintings
Joint statistics at the same scale
study of drawings

References


A Color conversion

The following is the conversion matrix from LMS to \(\alpha\beta\) [Ruderman et al. 1999].

\[
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix} = \begin{bmatrix}
\sqrt{\frac{2}{3}} & 0 & 0 \\
0 & \frac{1}{\sqrt{6}} & 0 \\
0 & 0 & \frac{1}{\sqrt{3}}
\end{bmatrix} \begin{bmatrix}
1 & 1 & 1 \\
1 & 1 & -2 \\
1 & -1 & 0
\end{bmatrix} \begin{bmatrix}
\log L \\
\log M \\
\log S
\end{bmatrix}
\]

The conversion matrices from RGB to LMS, from RGB to XYZ and from XYZ to LMS are also shown in [Reinhard et al. 2001a].

\[
\begin{bmatrix}
L \\
M \\
S
\end{bmatrix} = \begin{bmatrix}
0.3811 & 0.5783 & 0.0402 \\
0.1967 & 0.7244 & 0.0782 \\
0.0241 & 0.1288 & 0.8444
\end{bmatrix} \begin{bmatrix}
R \\
G \\
B
\end{bmatrix}
\]

\[
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix} = \begin{bmatrix}
0.5141 & 0.3239 & 0.1604 \\
0.2651 & 0.6702 & 0.0641 \\
0.0241 & 0.1228 & 0.8444
\end{bmatrix} \begin{bmatrix}
R \\
G \\
B
\end{bmatrix}
\]

\[
\begin{bmatrix}
L \\
M \\
S
\end{bmatrix} = \begin{bmatrix}
0.3897 & 0.6890 & -0.0787 \\
-0.2298 & 1.1834 & 0.0464 \\
0.0000 & 0.0000 & 1.0000
\end{bmatrix} \begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix}
\]
Figure 15: Pissarro’s *Mathurins* original, with Manet histogram and with Signac histogram

Figure 16: Signac’s *Dining* original, with Manet RGB histogram, with Manet lεβ histogram
Figure 17: A natural image, with Manet $\alpha\beta$ histogram, with Signac $\alpha\beta$ histogram and with Adams $\alpha\beta$ histogram
Figure 4: 3D plot of the Fourier transform of Manet's *Emile* and *Olympia* and of Monet's *Hay* and *Poplars*. 