Statistical Analysis and Transfer of Coarse-Grain Pictorial Style

by

Soonmin Bae

B.S. Computer Science, Korea Advanced Institute of Science and Technology (2003)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2005

Author .....
Department of Electrical Engineering and Computer Science
June 27, 2005

Certified by .....
Frédo Durand
Assistant Professor
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Statistical Analysis and Transfer of Coarse-Grain Pictorial Style

by

Soonmin Bae

Submitted to the Department of Electrical Engineering and Computer Science
on June 27, 2005, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

We show that image statistics can be used to analyze and transfer simple notions of pictorial style of paintings and photographs. We characterize the frequency content of pictorial styles, such as multi-scale, spatial variations, anisotropy properties, using a multi-scale and oriented decomposition, the steerable pyramid. We show that the average of the absolute steerable coefficients as a function of scale characterizes simple notions of "look" or style. We extend this approach to account for image non-stationarity, that is, we capture and transfer the spatial variations of multi-scale content. In addition, we measure the standard deviation of the steerable coefficients across orientation, which characterizes image anisotropy and permits analysis and transfer of oriented structures.

We focus on the statistical features that can be transferred. Since we couple analysis and transfer, our statistical model and transfer tools are consistent with the visual effect of pictorial styles. For this reason, our technique leads to more intuitive manipulation and interpolation of pictorial styles. In addition, our statistical model can be used to classify and retrieve images by style.

Thesis Supervisor: Frédo Durand
Title: Assistant Professor

Acknowledgments

First, I would like to thank God for His unfailing love and grace. Without his help, this thesis would not have been possible.

Thanks to my advisor Professor Frédo Durand, not only for his scholarly guidance and keen insight, but also for his patient and friendly support. I am grateful to him for letting me work on this exciting research topic. He has been a true mentor and been enthusiastic throughout the work on this thesis.

I also appreciate the encouragement and help of the whole Computer Graphics Group at MIT. They deserve my deepest thanks.

I thank my family and my friends. Their support and encouragement has been invaluable for me.

Lastly, I would like to thank the Samsung Lee Kun Hee Scholarship Foundation and the National Science Foundation for their financial support.

Contents

1 Introduction **6**

1.1 Overview of the Approach 8

1.2 Thesis Overview 10

2 Background **11**

2.1 Art Movement 11

2.2 The Human Visual System and Visual Information Encoding 16

2.3 Multi-Scale and Oriented Image Representation 20

2.4 Machine Learning Techniques for Feature Extraction 23

3 Image Statistics **26**

3.1 Marginal Statistics 27

3.2 Joint Statistics 28

3.3 Non-stationary Statistics 29

3.4 Applications 31

4 Previous Work **32**

4.1 Statistical Analysis on Sound and Music 32

4.2 Image Retrieval and Classification 33

4.3 Pictorial style Retrieval and Classification 34

4.4 Style Transfer 36

5 Image Features **38**

5.1 Multi-scale Amplitude 38

5.2 Non-stationarity 48
5.3 Anisotropy 54
5.4 High and Low Residual features 58
5.5 Summary 58

6 Style Analysis **59**
6.1 Principal Component Analysis (PCA) 60
6.1.1 Multi-scale Analysis 61
6.1.2 Non-stationarity Analysis 64
6.1.3 Anisotropy Analysis 68
6.1.4 Correlation and Dependency Analysis 72
6.2 Linear Discriminant Analysis (LDA) 74
6.3 Summary 78

7 Style Transfer **79**
7.1 Multi-scale Content Transfer 80
7.2 Non-stationarity Transfer 83
7.3 Anisotropy Transfer 87
7.4 Comparison 89

8 Discussion **94**

Chapter 1

Introduction

Our goal is to analyze and transfer coarse-grain aspects of pictorial style. We characterize them using quantitative features and show how such features are relevant to simple notions of pictorial styles. In particular, we are interested in features that can be transferred between images. We seek to separate style and content and build a statistical model of pictorial styles, with applications such as style transfer between images and image retrieval based on styles of images.

The goal of computer graphics has been a faithful illustration of the real world. Recently, along with the new field of non-photorealistic rendering, interest in broadening rendering styles has increased. Many researchers have developed tools for design and reproduction of pictorial styles. However, in general, pictorial styles have been explored implicitly, without explicit characterization. In this work, we seek to characterize pictorial styles of images quantitatively.

We are interested in coarse-grain notions of style. We restrict the aspects of style to the very noticeable visual qualities that humans detect before they scrutinize images closely. For example, the difference between a Pointillist painting and a Renaissance painting is obvious since the former uses small dots whereas the latter uses smooth brush strokes. We aim to parameterize and transfer such visible differences of the styles.

We focus on features that we can measure and analyze quantitatively. Subjective feeling about the color schemes and contents of images is not discussed in this thesis.

Certainly, we do not pursue the characterization of the original and creative skills of artistic geniuses. Moreover, labeling pictorial styles by artist or time periods is not our main concern. Instead, we are interested in the visual similarities and differences of pictorial styles.

The main originality of this work is the development of statistical features that permit *both* the characterization and transfer of pictorial styles. We decouple the visual effect of pictorial styles from contents of images and statistically characterize them. We seek to change and transfer the style without changing the original content. We believe that style transfer is a powerful evaluation tool that validates the visual relevance of our features.

Our technique is useful for image retrieval, classification, and manipulation, while providing easy and intuitive user interfaces. Our tool facilitates achieving images with the visual appearance a user seeks.

Motivation Users have been empowered by a large amount of available digital images and by powerful hardware and software. They make images, share images with their friends and families, and sometimes make the images public. Moreover, they modify and enhance images using image editing software. In addition, users can search images using content-based image retrieval systems.

However, it is not always straightforward to obtain images with a desired look and feel. Existing hardware and software tools tend to be complicated. They require a large amount of skill and effort. Casual users need a more intuitive way to obtain an image with the style they want. One way to achieve an image with a given look is to take advantage of model images that already have the desired characteristics and *transfer* such a look. This method is useful, because it does not require refined artistic skills. Using a model image, users can search for images with the desirable look. Alternatively, users can transfer visual qualities of the model image to their own image.

Background We analyze the styles of notable artists and photographers, which provide good insights for the characterization of pictorial styles. In this thesis, we compare our notions of style with those of art history, although we do not seek a strict match.

In addition, our approach is inspired by knowledge about the low-level human visual system. An understanding of the human visual system helps us connect elusive notions of style with computational features. Ideally, we try to characterize styles using the features that humans use to appreciate images.

It must be understood that the goal of this project is not to reproduce the stylistic notions in art history or the inference mechanism in the human visual system. We focus on stylistic features that are content-independent, transferrable or parameterizable, and free of high-level visual processing.

1.1 Overview of the Approach

In this work, we focus on style *analysis* and *transfer*. We use image-processing and statistical tools to extract features characterizing pictorial styles. We validate the correlation of the features with pictorial styles using machine learning techniques. Transferring style from one picture to another further assesses the visual relevance of the features.

To build our descriptive model of pictorial styles, we use a *data-driven inductive method*: we propose assumptions on the visual regularities and variations in visual arts, and validate the assumptions by finding statistical features. The statistical features should be relevant to the visual regularities and variations, and generate the visual appearance in the paintings and photographs. To validate the statistical model we take advantage of machine learning techniques, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). These two tools are used to examine whether the features represent the styles consistently.

We perform our analysis and transfer based on frequency content. For the statistical model and representation of styles, we use a multi-scale and oriented image

Artist	Nationality	Year	Art movement
Beckmann	German	1884-1950	Expressionism
Bohrod	American	1907-1992	Surrealism
Chagall	Russian	1887-1985	Expressionism
Degas	French	1834-1917	Impressionism
Kandinsky	Russian	1866-1944	Expressionism
Manet	French	1832-1883	Impressionism
Monet	French	1840-1926	Impressionism
Rembrandt	Dutch	1606-1669	Baroque
Renoir	French	1841-1919	Impressionism
Rubens	Italian	1577-1640	Baroque
Signac	French	1863-1935	Pointillism
Van Gogh	Dutch	1853-1890	Expressionism

Table 1.1: The list of artists we analyze in this project

representation. Such a representation has decomposition properties that resemble the low-level stages of the human visual system. Among multi-scale and -oriented image representation techniques, we choose the steerable pyramid [23, 61, 62]. The steerable pyramid has good reconstruction properties: it is free of aliasing and uses a smooth reconstruction basis. We perform image-processing and statistical analysis in the CIELab color space [46]. The CIELab color space is perceptually uniform and has uncorrelated color axes that correspond to the color opponents in the visual system.

Our Image Set We selected twelve artists from the 1600s to the 1900s, from Mark Harden's Artchive [1]. We also use casual photographs as references, as well as photographs from masters such as Ansel Adams, Doisneau, Steve McCurry and famous photos from *National Geographic*. The artists are selected to span a wide range of style variation: some artists have very distinctive styles and others might have somewhat similar styles. Table 1.1 lists the artists we study in this work. They have various nationalities, were of different periods, and followed different art movements. For faithful comparison, the images are all down-sampled so that the minimum dimension is 512 pixels.

1.2 Thesis Overview

This thesis addresses the statistical analysis and transfer of pictorial style. We seek to find a relevant feature set that characterizes style similarities and differences. Chapter 2 introduces some background information that is useful for this thesis. Chapter 3 demonstrates previous work of image statistics and our relevant findings. Chapter 4 presents previous work including statistical analysis of sound and music, image classification and retrieval, pictorial style classification and retrieval, and style transfer. Chapter 5 discusses the set of features we use. Chapter 6 validates how those features are relevant to pictorial styles using machine learning techniques. Chapter 7 shows the results of the style transfer. Chapter 8 summarizes our work and discusses our future work.

Chapter 2

Background

Art history and our simple notions of pictorial styles have different perspectives as well as common viewpoints. We want to discuss these discrepancies and congruencies. Research on image statistics and the mechanics of the human visual system provides important tools for the analysis and transfer of these regularities and variations of pictorial styles. Our work is facilitated by tools in computer vision and image processing. We use machine learning techniques to validate our analysis of these statistical features.

The first section in this chapter covers an informal discussion of stylistic trends that prevailed from the 1600s to the 1900s. The second section introduces the basic mechanism in the human visual system (HVS), which covers the eye and a part of the human brain that is responsible for processing of the visual image. The next section discusses various image representations related to the human visual system. The last section in this chapter reviews machine learning techniques we use to validate our statistical model.

2.1 Art Movement

We selected twelve artists from the 1600s to the 1900s, where each movement has distinct characteristics. The rest of this section describes several categories used as illustrations in this thesis such as Baroque, Impressionism, Pointillism, and Expres-

sionism and discusses the issues between the conventional labeling and our simple notion of styles. The purpose of this section is not to provide an introduction to art history. Moreover, the goal of this thesis is not to find a style model that matches notions of pictorial styles of art history. We refer interested readers to Gombrich's *The Story of Art* [24].

We focus on characterizing the *visual* similarities and differences between pictorial styles of paintings, rather than labeling styles by movements or by time periods. Pictorial styles are usually characterized based on aspects such as time periods, nationality, subject, and the intent of artists. However, neither the definition of individual styles nor the categorization of the artists is unanimous even among critics. As artists explore original pictorial styles, two different artists in the same category do not necessarily share the similar visual styles. In addition, the pictorial style of an artist might vary significantly during his or her career, which is in the case for painters such as Picasso or Kandinsky. This ambiguity and complexity of the notion of style makes the labeling of data for our study a difficult issue.

Figure 2-1: Rubens' *Daniel in the Lion's Den* and Rembrandt's *The Feast of Belshazzar*

The work of Rubens and Rembrandt belongs to the Baroque movement that emerged in Europe around 1600. At that time, art was used to present religious subjects or portraits in a very realistic way. Paintings played the role of photographs, which did not exist at that time. Objects were drawn with details and each painting

had a specific subject and real objects to depict. To present the subject and the main object, the degree of details was carefully arranged. (See Fig. 2-1.)

Between Baroque movement and Impressionism, there exists several movements such as Neoclassicism and Romanticism. Neoclassicism sought to revive the ideals of ancient Greek and Roman art, and Romanticism painted emotions in a romantic manner. Since our data set does not have any artist categorized as these movements, we do not describe them in detail.

Considering subject matter and pictorial style, more diverse exploration had started around the 1800s, and various art movements were founded in the 1800s and 1900s, which included Post-impressionism, Pointillism, Fauvism, Expressionism, Cubism, and Surrealism. Above all, the aim of painting shifted from an imitation of nature to an expression of the artist's impression and emotion. For example, impressionists focused on visual perception and aesthetics, expressing what they perceived. Expressionists explored ways of using paint and brushes to express what they imagined or what they felt.

Impressionism was developed in France in the late 1800s and early 1900s. The Impressionists include Degas, Manet, Monet, Pissarro, Renoir, and Sisley. Artists turned to drawing daily-life subjects instead of focusing on the biblical stories and portraying noble people. They tried to capture an immediate visual impression of their subjects by using color rather than well delineated figures. They applied colors separately onto the canvas in small dabs and dashes instead of mixing pigments. They focused on an overall effect rather than on details using much more visible strokes.

Pointillists, such as Seurat and Signac, used point-shaped brush strokes instead of traditional brush strokes to achieve the overall effect by blending the colors of the dots to form a picture in the hope of creating more vivid colors. The similarities and differences between Impressionism and Pointillism are presented in Monet's painting and Signac's painting in Figure 2-2. Both depict natural scenes using distinct brush strokes and bright colors, while their ways of using colors and small dots distinguish each other.

A comparison between Figure 2-2 and 2-1 reveals that various degrees of differences

Figure 2-2: Monet's *Poplars along the River Epte, Autumn*, Signac's *The Papal Palace, Avignon*, Degas's *The Cotton Exchange in New Orleans*, and Renoir's *Portrait of Bazille*

exist between pictorial styles of Baroque artists and Impressionists. Although Degas and Renoir are Impressionists, their paintings in Figure 2-2 are very realistic and more similar to the paintings of Baroque artists in Figure 2-1 than to Monet's painting in Figure 2-2. Although Signac painted in the similar time period as Degas and Renoir did, the paintings by Degas and Renoir is very different from that by Signac in Figure 2-2, but rather similar to Baroque artists, who had lived two hundred years before they did. This shows that categorization by time period and movement can be different from that by visual difference and similarity.

Figure 2-3: Kandinsky's *Composition VIII* and *With Black Arch* and Chagall's *Maternity*

Expressionists explored the expressive potential of paintings and used more vivid colors and distortion for expressing the artist's ideas on emotion. The objects were not necessarily things in the real world and the objects were not depicted as they looked in real life. The arrangement of objects was explored and the color and shape was carefully designed. In a broad sense, Expressionists include Chagall, Kandinsky, and Beckmann. Although they were not formally expressionists, the work of Gauguin and van Gogh had a critical impact on this movement. In addition, Kandinsky is considered to be a founder of abstract art. Since individual artist explore different pictorial styles, it is difficult to categorize each artist. Figure 2-3 shows some examples. The visual difference between two paintings by Kandinsky is larger than the difference between Kandinsky's *With Black Arch* and Chagall's *Maternity*.

We have shown that labeling pictorial styles is complex and can be ambiguous. Conventional labeling of pictorial styles such as time periods and artists can be different from labeling based on visual differences and similarities. In our work, we study pictorial styles using coarse-grain notions of styles based on visual aspects such as large-scale contrast, spatial distribution of details, and the use of texture, lines, and edges.

2.2 The Human Visual System and Visual Information Encoding

Studies of human visual perception provide insights for computer vision and computer graphics. Our work benefits from these studies. While considering image features for quantitative style characterization, we take into account their relevance to properties of the human visual system such as color perception, contrast sensitivity and multi-scale and oriented encoding schemes. In this section, we describe how our choices of image processing techniques are linked to such properties of the human visual system.

The first phase of the human visual perception occurs in the retina of the human eye. Figure 2-4(a) illustrates the structure of the human eye. The retina, the thin

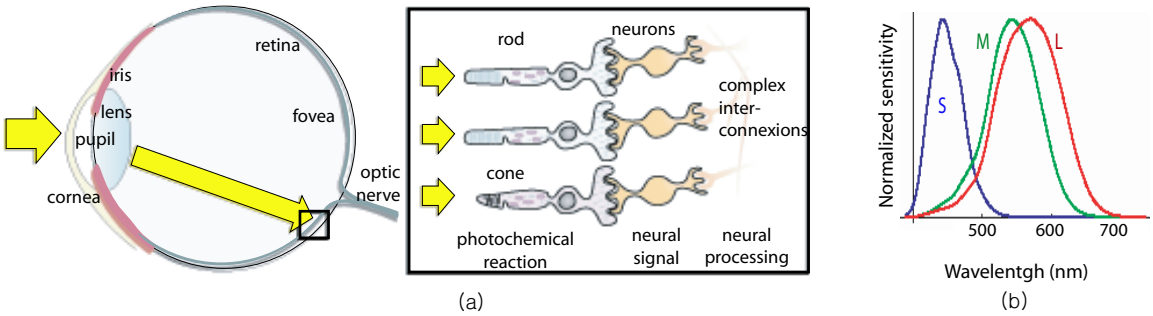
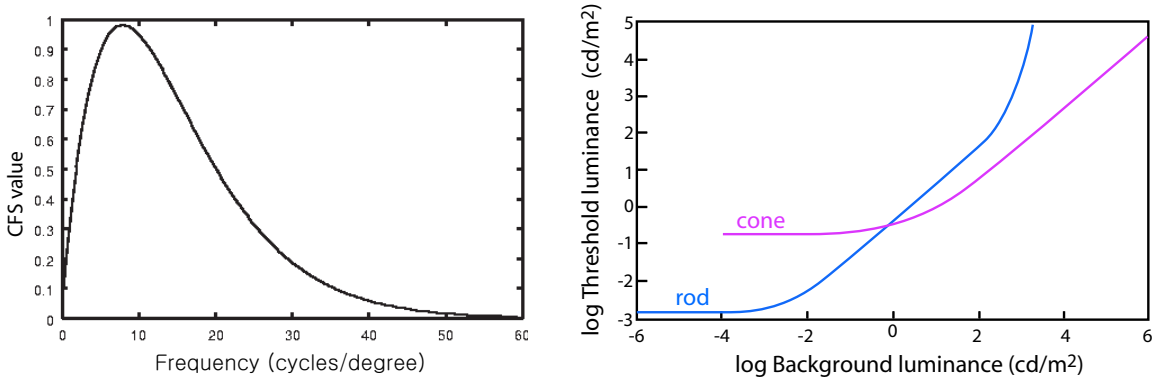


Figure 2-4: (a) The human eye. The structure of the retina is zoomed in. (b) The wavelengths to which the three different types of cones are sensitive. (Image courtesy of Frédo Durand)

layer at the back of the eye, contains the photoreceptors and has retinal ganglion cells on its surface.

The retina has two types of photoreceptors, the cones and the rods. Rods are more sensitive to light than cones [66]. On the other hand, color perception is mainly performed according to the responses of the cones. There are three types of cone photoreceptors, called L cones, M cones, and S cones, distinguished by different wavelength sensitivity. Figure 2-4(b) shows the wavelength sensitivity of these cones: L cones are most sensitive to the long wavelength, M cones to the middle wavelength, and S cones to the short wavelength. The relative spectral power distribution of the incident light plays an important role in the discrimination of the color appearance of light.

The perception of color is determined by the relative magnitude of the signals from these cones, as described by the trichromatic theory of color vision [66]. According to the color opponent theory, the responses of the three cones are then remapped into three neural channels: one achromatic channel and two chromatic channels, that is, black-white, red-green, and blue-yellow. The black-white channel sums responses from all three cones. The ratio of response of the L cones in contrast to that of the M cones determines the red-green channel. The blue-yellow channel encodes the ratio of the response of S cones in contrast to the sum of L and M cones' responses.



(a) (b)

Figure 2-5: (a) The contrast sensitivity function. (b) The threshold-versus-intensity curve in the log-log domain (Image courtesy of Frédo Durand)

For luminance, the human visual system has been shown to be more sensitive to local contrast than to absolute luminance. The contrast sensitivity function illustrates that contrast perception shows different sensitivities across spatial frequencies in Fig. 2-5(a). In addition, it is known that just-noticeable contrast perception is consistent across most of the useful range of luminance: the Weber-Fechner law [18, 69] states that the logarithm of intensity threshold increases linearly as the logarithm of background intensity. (See Fig. 2-5(b).) In addition to the Weber-Fechner law, there have been a number of studies on the perception of luminance and contrast. While these issues are related, they focus on different aspects of light perception. In particular, the Weber-Fechner law characterizes near-threshold perception, that is, at which level of difference humans can see the difference between two colors. In contrast, other studies deal with supra-threshold perception.

For example, Stevens related subject brightness to intensity, according to the following power law [66].

$$B = k\, I^{\alpha},$$

where B is subjective brightness, I is intensity, and α is 0.33 in the case of brightness.

Munsell sought to identify equal perceived color differences [16]. The Munsell

color appearance model has three attributes: Value(V), Hue(H), and Chroma(C) [16]. According to the Munsell Value scale, 10 corresponds to white and 0 to black. The Munsell Hue circle is divided into five principal hues: red, yellow, green, blue, and purple. The Munsell Chroma scale has a maximum depending on the Munsell Value and hue: yellow at its highest chroma has a much lighter value than does blue at its highest chroma, since the Munsell color space is not a cylinder nor a sphere. The CIELab color space is similar to the Munsell color model and designed to be perceptually uniform: the Euclidean distance in the 3D Lab space corresponds to perceptual color difference. However, this is true only for small distances. These properties of the CIELab color space are favorable for image analysis and transfer. For the same reason, we perform multi-scale decompositions in the CIELab color space.

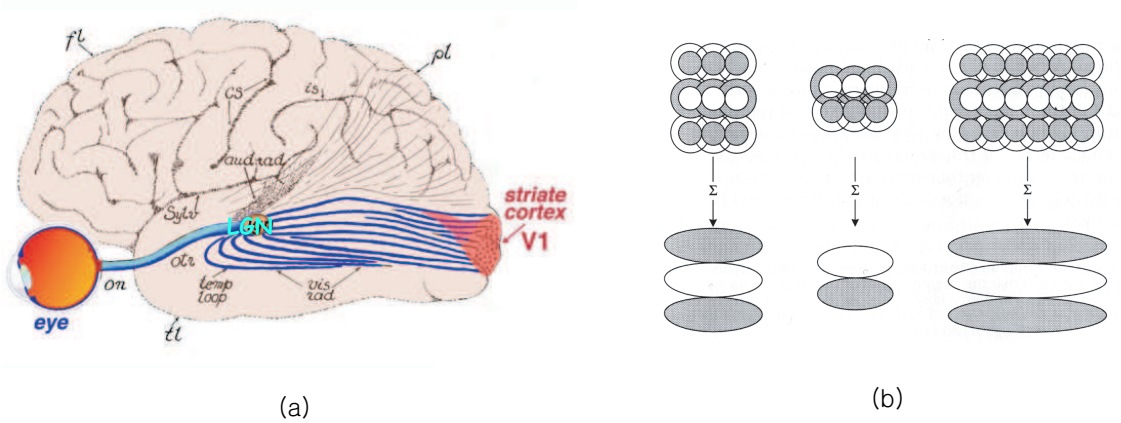


Figure 2-6: The human visual system. (a) Light goes from eye to the lateral geniculate nucleus (LGN) and then to primary visual cortex (V1) from Polyak(1957) [48]. (b) The illustration of the receptive fields of three hypothetical neurons. Each has different degrees of orientation selectivity from Wandell(1995) [66].

After the visual information passes through the photoreceptors of the retina, it is processed through the center-surround receptive fields of the retinal ganglion cells and the lateral geniculate nucleus (LGN) [66]. The response properties of these receptive fields have been hypothesized to be consistent with those of linear-systems

[66]. Once the visual information is processed in the retina and LGN, it is passed to the primary visual cortex, called V1. (See Fig. 2-6(a).) Studies on the human visual system propose that V1 encodes visual information in terms of spatial frequencies and orientation. Accordingly, the receptive fields of simple V1 cells have been shown to have preferred orientation and scale [66]. Figure 2-6(b) illustrates receptive fields with different degrees of orientation selectivity. Multi-scale image representations such as the 2D Gabor wavelet have been found to be consistent with the structure of simple cell receptive field properties in the visual cortex [7–9].

Thus, multi-scale and oriented image representations are consistent with measurements of the receptive field properties in the primary visual cortex. For this reason, among image processing tools, we choose steerable pyramids [19, 61, 62] to compute image statistics, since the steerable pyramids provide multi-scale and oriented decomposition, which we describe in the next section.

2.3 Multi-Scale and Oriented Image Representation

We perform our analysis and transfer using a multi-scale and oriented image representation called steerable pyramid. The steerable pyramid is suitable for our work in terms of their decomposition and reconstruction properties. In this section, we describe these properties of the steerable pyramid compared to those of the Fourier transform and wavelet transforms.

The Fourier transform is a classic tool in image processing. It is a simple linear basis transform where an image is represented as a sum of different frequencies. Unfortunately, the Fourier transform is not suitable for a signal localized in space because of its infinite support basis. For this reason, wavelet transforms have provided a powerful alternative. An additional advantage of wavelet transforms is that the coefficients of wavelet subbands are suitable for a sparse encoding. These properties have been used for image compression, restoration, and analysis and synthesis [4, 59].

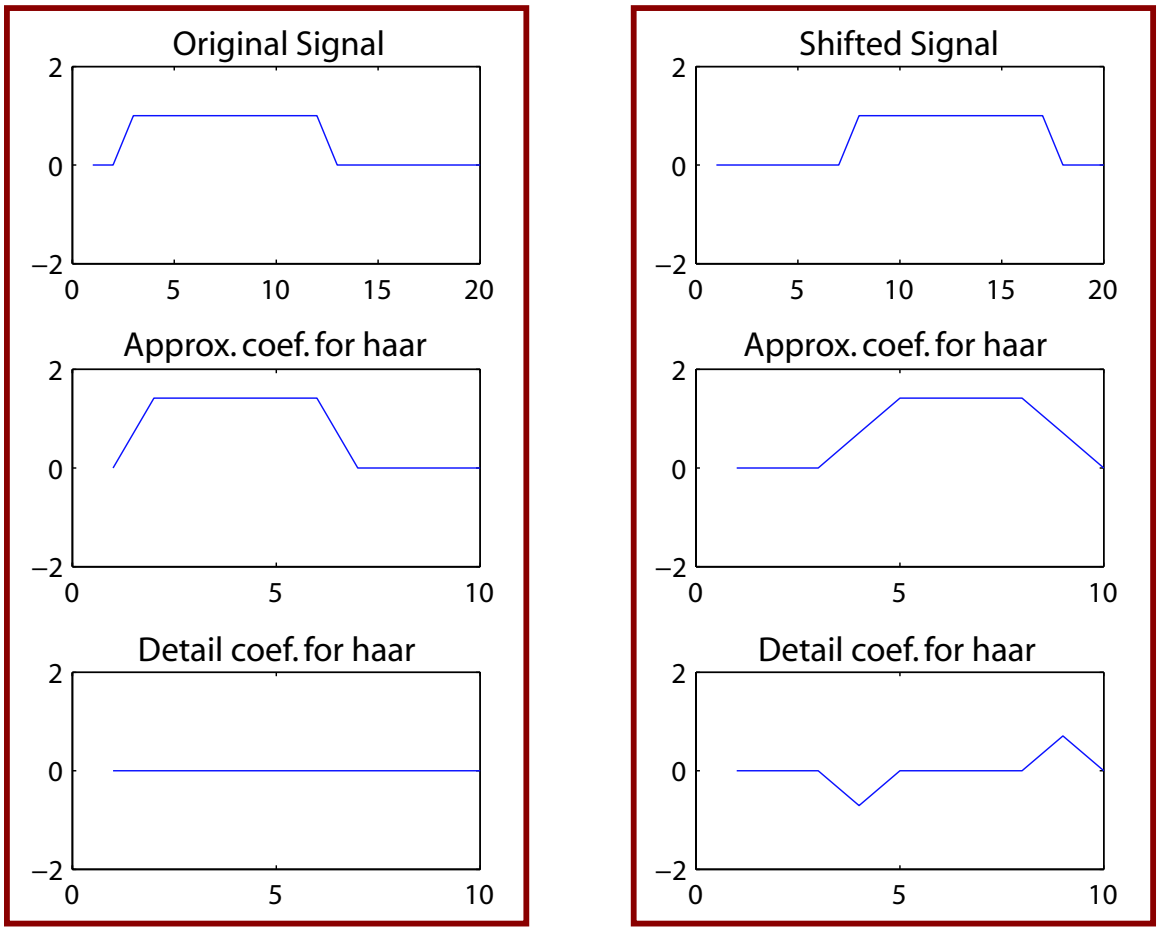
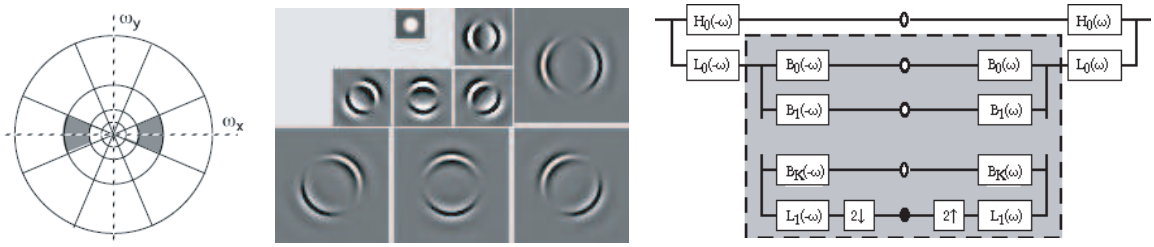


Figure 2-7: Orthnormal wavelet representation does not have shift-invariance. We demonstrate such property using a Haar wavelet.

However, orthonormal wavelet representations are not band-limited and suffer from a lack of translation-invariance, which causes artifacts in applications like texture analysis and synthesis [62]. Figure 2-7 shows an example of the lack of shift-invariance.

The steerable pyramid [23, 61, 62] is a good substitute for the wavelet transform. Unlike the orthogonal separable wavelet, the steerable pyramid is translation-invariant and rotation-invariant, which lead to good reconstruction properties: it is free of aliasing and uses a smooth reconstruction basis. However, a shortcoming of this representation is overcompleteness.



(a) (b) (c)

Figure 2-8: (a) A spectral representation of the steerable pyramid with 4 orientations. (Images from Simoncelli and Freeman [61] (b) A 2-level decomposition of an image of a white disk using the steerable pyramid with 4 orientations. (c) The block diagram for the decomposition and reconstruction. (Images from http://www.cns.nyu.edu/∼eero/STEERPYR/)

In addition to localization in space and frequency, the steerable pyramid provides subbands localized in orientation. (See Fig. 2-8 (a) and (b).) The steerable pyramid is constructed as follow. First, the image is decomposed into low and highpass subbands. Then the lowpass subband is divided into a set of oriented bandpass subbands and a lower-pass subband. The lower-pass subband is subsampled by a factor of 2 in the X and Y directions. This process is done recursively to construct a pyramid. (See Fig. 2-8 (c).) In summary, each subband is tuned to be localized in scale and orientation.

The multi-scale and oriented decomposition properties of the subbands of the steerable pyramid resemble the simple cell receptive fields in the human visual system. On account of its oriented multi-scale decomposition and good reconstruction

properties, we choose the steerable pyramid for our analysis and transfer.

2.4 Machine Learning Techniques for Feature Extraction

The extraction of relevant and useful information from a large set of high-dimensional data is not trivial. The sheer amount of data makes it difficult to visualize and understand. Moreover, the features might be neither independent nor uncorrelated, and the examination of data along an individual dimension does not reveal the underlying trends. Machine learning techniques can be used to reduce the number of dimensions for the purpose of analysis. Machine learning techniques are suitable to extract the meaningful projection of features, i.e. to reveal correlations and dependencies.

Before extracting a projection of features, we need to choose a set of features. This is called feature selection. In this phase, we should be inclusive and not discard meaningful aspects or dimensions of data. Once we have selected features, we can use machine learning techniques to transform the set of features into a more interpretable and compact feature set that specifies meaningful correlations and dependencies. Such feature transform techniques include Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Independent Component Analysis (ICA).

PCA and ICA are *unsupervised* learning techniques, and LDA is a *supervised* learning technique. The basic difference between supervised learning and unsupervised learning is whether labels are given or not in the training data. In the supervised learning, input-output pairs are given during a training phase. The goal of supervised learning is to find a function that generalizes the mapping between inputs and outputs. In contrast, outputs are not given for unsupervised learning. The unsupervised learning methods try to find suitable representations of inputs, which possibly generalize underlying structures of inputs. In our case, we use both learning without labels and with labels.

PCA finds the directions in the feature space along which data points are well-

spread. PCA uses second order statistics to decorrelate features. The principal directions of PCA are orthogonal to one another, and projections of data onto them are linearly decorrelated. With the decorrelated features, we can analyze additional aspects of data. However, imposing orthogonality might cause some underlying structures of data to be lost, and the inherent scaling indeterminacy can mislead the principal axis of PCA. That is, if the relative scaling of features is arbitrary, relevant features might be scaled down and irrelevant features are not scaled down properly.

Like PCA, ICA decorrelates features. However, ICA assumes that features are statistically independent, not merely decorrelated. Therefore, it is not appropriate for data that are not independent.

For LDA, we assign labels to the data points of two different classes, then LDA computes a linear classifier that separates data points of different classes by a hyperplane, finding the direction and offset of the hyperplane. LDA assumes that the densities $p(\vec{x}$ is in $class$ 1) and $p(\vec{x}$ is in $class$ 2) are both normally distributed, with identical covariance, but possibly different means, μ_1 and μ_2. $P(which\ class$ is $\vec{x}$ in) is then determined by $\vec{x} \cdot \vec{w}$, where

$$\vec{w} = \Sigma^{-1}(\vec{\mu}_2 - \vec{\mu}_1).$$

Thus LDA maximizes the ratio of the distance between groups to the distance within group. When labels are available, LDA has the advantage over PCA that it can extract the relative scaling of features to achieve optimal linear classification.

When we use these machine learning techniques to find a model that approximates and generalizes trends in data, we should be careful not to choose a model that overfits the training data. When a classifier is overfitting the training data, it might not able to give good predictions on new unseen data. Overfitting occurs when the learner tries to fit very specific features, such as random noise, which are not relevant to regularities. A complex model may increase the performance on the training data, while making the performance on unseen data worse.

We use PCA and LDA to validate our statistical model. These two tools examine

whether the features represent the styles consistently. When we use machine learning tools, we restrict our generalization to linear modeling. Linear modeling is usually simple enough to prevent overfitting of training data. This property of linear modeling is necessary, since we have a limited number of images corresponding to each style. A complex model is likely to fit random noise or content, not just the regularities of pictorial styles, leading to overfitting.

Chapter 3

Image Statistics

Statistical characterization has been applied to measure regularities and differences among so-called "natural" images. The striking insight of this new field is that not all random 2D signals look like natural images. This has led to the development of tools to statistically characterize what is specific to natural images, and how to distinguish classes of natural images. We use the comparable statistical characterizations to analyze and transfer pictorial styles. In this chapter, we review statistical features studied in the fields of natural image statistics and show corresponding statistical features of paintings and artistic photographs.

Many researchers in various fields including human vision, machine vision, and image processing have explored the statistical characterizations of natural images and found statistical regularities among them [20, 28, 29, 53–55, 58, 60]. Image analysis has used these regularities, and the variation of the statistical properties has been applied to image classification and synthesis [10, 11, 25, 50, 64]. This chapter shows that we observe those regularities present in paintings and artistic photographs. In this thesis, we propose that certain properties of image statistics can be used to analyze and transfer the similarities and differences of pictorial styles.

This chapter describes previous work in natural image statistics and our corresponding results for paintings and artistic photographs. The first three sections discuss marginal statistics, joint statistics, and non-stationary statics. The last section introduces applications of these tools.

3.1 Marginal Statistics

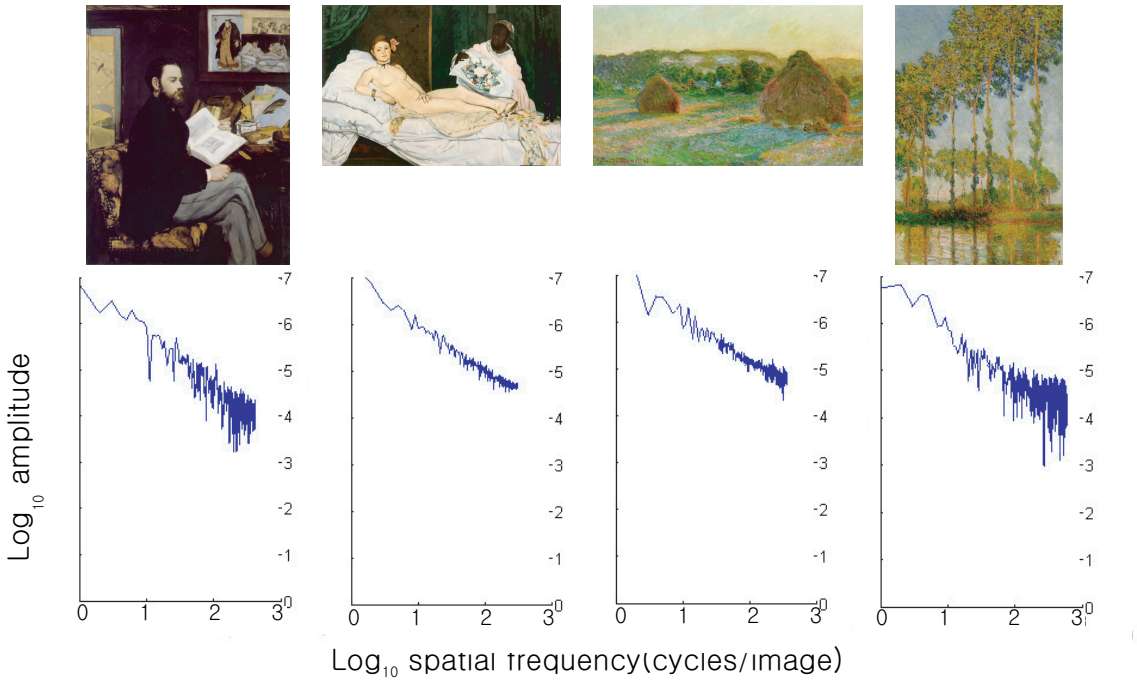


Figure 3-1: The average amplitude curves in the log-log domain. Their slopes show that average amplitude falls as "1/f".

Statistical characteristics of natural images are often measured in the frequency domain. Frequency analysis provides the nice property that the power spectrum is the Fourier transform of the autocorrelation function. Field investigated the two-dimensional amplitude spectra and found regularity among natural images [20], as the average amplitude falls as "$\frac{1}{f}$". This property is called scale invariance [52, 54, 55]. In Fourier terms, this implies that any spatial frequency band with the same octave bandwidth has an equal amount of energy. We observe that this statistical regularity is mostly shared by paintings. The average amplitude curves in Figure 3-1 show examples of scale invariance in paintings. In addition, we find statistical regularities can be represented in terms of an average of scaled coefficients of the steerable pyramids. The flat curve of Fig. 3-2(c) corresponds to the scale invariance. In this thesis, we will focus more on the differences of frequency content we found in the curves.

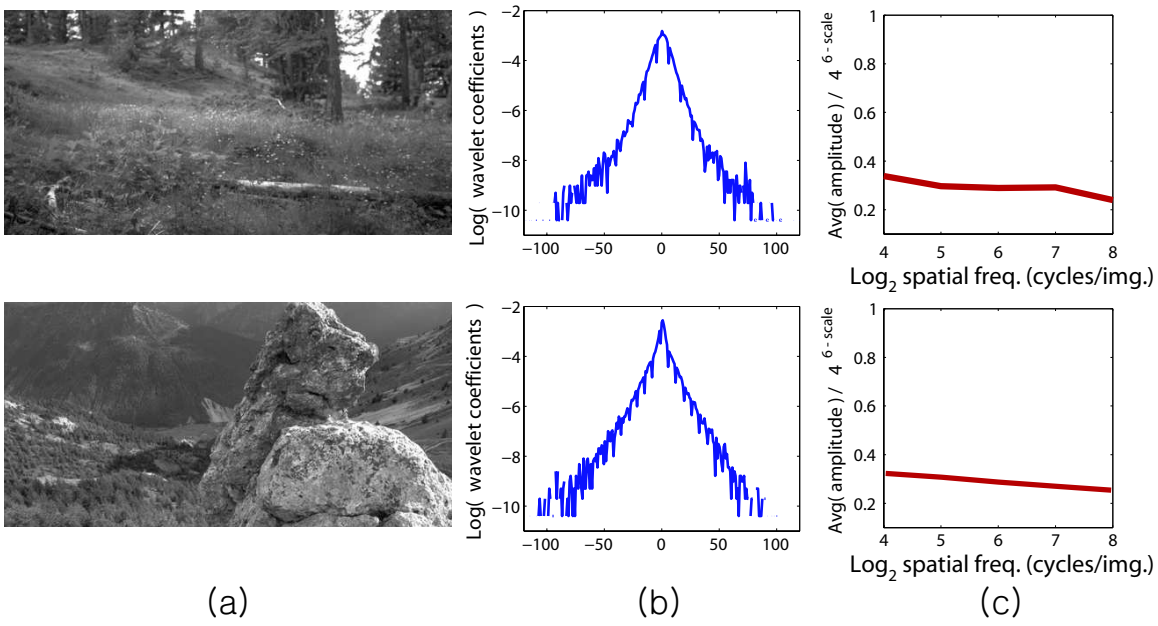


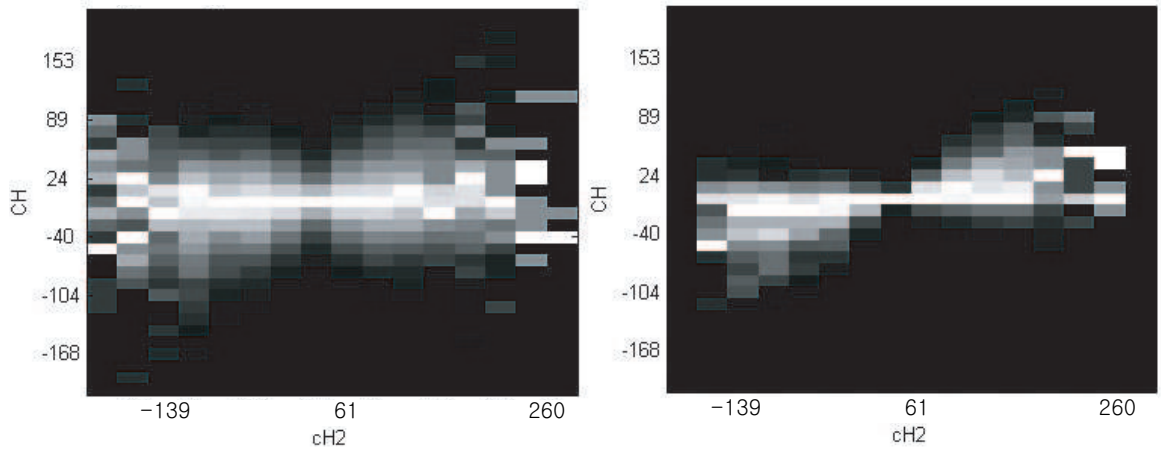
Figure 3-2: (a) Natural images. (b) The marginal distribution of Haar wavelet coefficients in the log domain. (c) The scaled average amplitude of steerable pyramid coefficients.

Regularities have also been found in other image features. Huang et al. [28, 29] discuss statistical regularities in the wavelet domain. Figure 3-2 shows some examples, such as non-Gaussian shape and a long tail in the distribution of coefficients.

In addition, the marginal distribution of the derivative, one calculated from the difference between two adjacent pixels, has been studied [11,28,29] and shown to have a high kurtosis with large tails and a peak at 0 in natural images.

3.2 Joint Statistics

Simoncelli [60] and Huang and Mumford [29] have explored the joint distribution of two adjacent pixels and have found a rough independence. In contrast, they have studied joint statistics in the wavelet domain and have found that coefficients of adjacent scales and orientations are not independent. That is, in natural images, higher-order statistics reveal high correlation between coefficients in the wavelet do-



(a) Signac (b) Manet

Figure 3-3: The conditional histogram H((C)|(P)) for horizontal wavelet coefficients. Bright parts mean high values.

main. We propose that the corresponding results for paintings are relevant to some characteristics of pictorial styles. Figure 3-3 illustrates that the joint statistics can capture the difference between Manet's and Signac's pictorial styles. The plots are conditional histograms of fine-scale coefficients given coarse-scale coefficients. Manet's conditional histogram shows stronger correlation than Signac's. In this thesis, we do not discuss pictorial styles in terms of joint statistics, but we will explore pictorial styles using joint statistics in our future work.

3.3 Non-stationary Statistics

So far, we have discussed statistical properties over the full image. However, image statistics have been shown to be non-stationary [64]: local statistical features vary with spatial location. Figure 3-4 shows an example where local frequency content varies according to its spatial location. The local spectral signatures are obtained by taking spectrum of each window, where we remove boundary effects by using a smooth windowing function. We use a hamming window in practice. The degree of

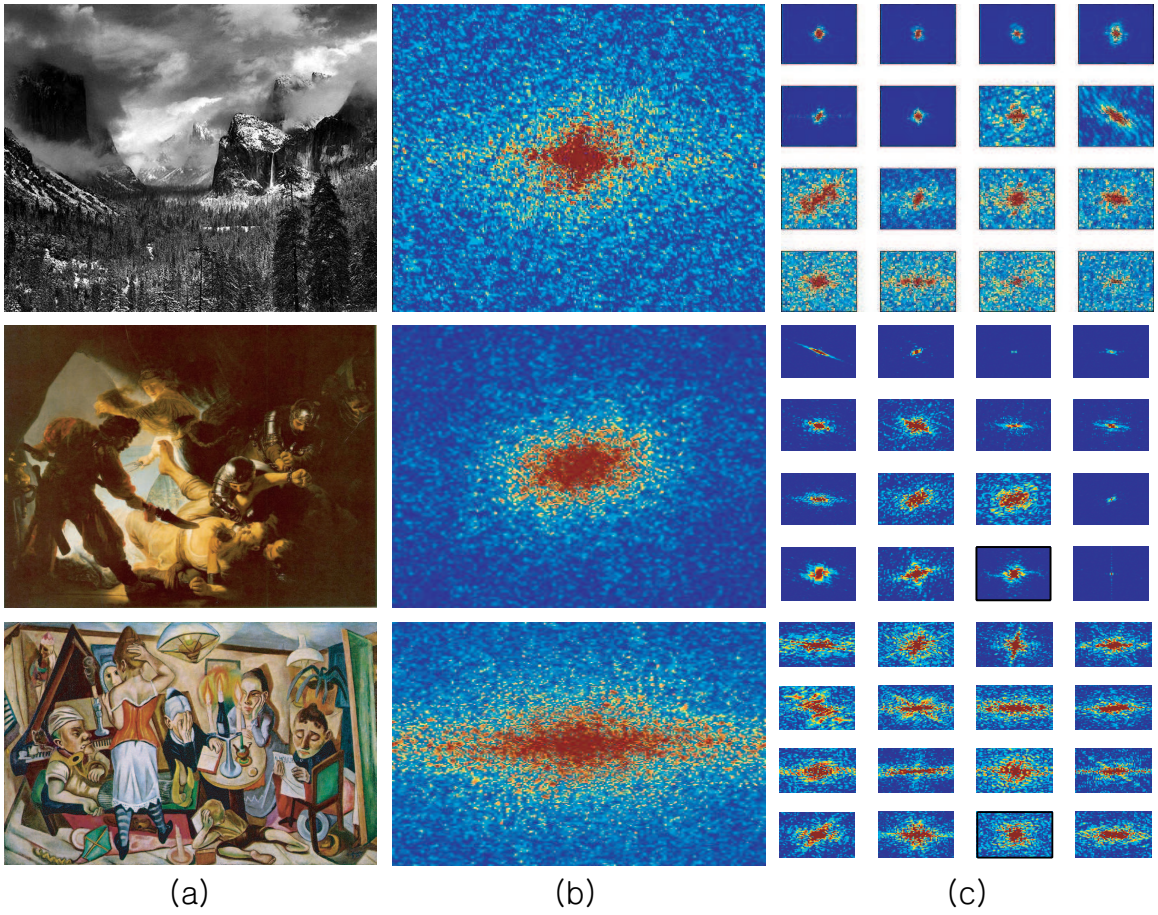


Figure 3-4: (a) The input image (b) The global spectral signatures. (c) The local spectral signatures. The color spectra represents the degree of the magnitudes. The color close to red means the high value and that close to blue means the low value.

the non-stationarity varies in different images. We will show that the variations in non-stationarity are relevant to pictorial styles.

3.4 Applications

We have discussed several statistical regularities in natural images, such as scale invariance, highly correlated joint statistics, and non-stationarity. These statistical regularities distinguish natural images from random two-dimensional signals such as white noise. The discriminative power of natural image statistics, in addition to the statistical regularities, have led to the exploration of the relationship between image statistics and the human visual system [2, 20]. The human visual system is hypothesized to efficiently encode the visual information using the statistical regularities present in natural images.

Statistical properties of natural images provide a prior on the set of relevant output images, which is fruitful in image processing applications such as denoising, compression, and texture analysis and synthesis. Marginal statistics and joint statistics in the wavelet domain have been used for image compression [4, 59] and denoising [49, 51]. Buccigrossi and Simoncelli [4] use the heavy-tailed and non-Gaussian marginal statistics and highly correlated magnitudes at adjacent scales, orientations, and spatial locations to build an image coder. Portilla et al. [51] use a statistical model of coefficients at adjacent positions and scale to remove noise.

While statistical regularities are used to analyze and encode natural images, the differences in image statistics can be used to classify and synthesize images [10, 11, 25, 50, 64]. We discuss previous work on image classification and synthesis in the next chapter. In this thesis, we propose that statistical properties can be used to analyze and transfer pictorial styles.

Chapter 4

Previous Work

In the previous chapter, we discussed statistical analysis of natural images. Statistical characterization has been applied to music, image classification and retrieval, and style transfer. Statistics are used to measure regularities and differences among images, musical genres, and pictorial styles. In this chapter, we review statistical features studied in the fields of musical genre and image retrieval and classification, and pictorial style retrieval, classification, and transfer.

4.1 Statistical Analysis on Sound and Music

Research on modeling music regularities and variations provides inspiration and insights on the scope and potential applications of analysis and transfer of pictorial styles. Widmer et al. provides very useful insights on possible goals and scope [72].

In general, music classifiers [43, 65, 70] classify songs into genres such as classical, rock, and so on. The success rate of classification is high, yet the level of discrimination is not fine enough to distinguish subtle differences between artists. This suggests that our expectation for pictorial style classifications should be bound to a coarse-grain level.

Most music classifiers use features that represent the spectral properties of music and their change in time. Some have also tried to characterize each instrument's timbre [13, 71]. This characterization of the timbres of each instrument permits sep-

aration of timbres from other perceptual aspects of music, such as dynamic, melody, tempo and so on. The decoupled timbre has been recombined with a set of perceptual parameters to produce a new music [31, 73]. As the decoupling of timbres ends in powerful music synthesis techniques, we propose that the separation of pictorial styles from content can lead to a direct manipulation and interpolation of pictorial styles.

Music analysis and synthesis has widely used spectral analysis to characterize musical styles or genres. However, not much research on pictorial styles has focused on the spectral aspects. We conduct our analysis and transfer in the frequency domain to characterize coarse-grain pictorial styles.

4.2 Image Retrieval and Classification

Since the early 1990s, image retrieval has drawn ever-increasing attention in the fields of computer vision, machine learning and databases in both the commercial and academic communities. Due to the rapid growth in the size of digital image collections, there is a growing need for efficient and easy-to-use image retrieval systems. A variety of content-based image retrieval (CBIR) systems have been developed [56, 63]. These CBIR systems retrieve images according to the contents of images instead of keywords. However, current CBIR systems do not search by semantic content, but by low-level visual features like color, texture, and shape [5, 21, 42, 47, 68, 74]. Some systems [21, 47] represent images using global features such as texture histogram, color histogram, and color layout of the whole image, whereas others use local features [42, 68]. Wang et al. [68] measure the similarity between images based on segmentation and region-based features, such as color, texture, shape, and location. They use supervised learning methods to classify the images into rough semantic classes, such as textures vs. non-textures and graph vs. photograph.

Some of CBIR systems use the similarity measure for clustering [74] or for classification [39]. In addition, there are CBIR systems that use machine learning techniques to build up high-level content, such as segmented objects [5] and semantic indexes [3, 40], for further clustering and classification. Chen et al. [74] use the same

features and similarity measure as Wang et al. [68], but cluster images into semantic groups and retrieve semantic clusters instead of a set of images. Carson et al. [5] connect keywords and image features. Barnard and Forsyth [3] establish a hierarchical clustering model linking image features to semantic information to organize image database. Likewise, Li and Wang [40] aim at automatic linguistic indexing of images.

While some of the CBIR systems use statistical models, they do not provide a model that is intuitively modifiable. We want to build a statistical model that is transferrable and modifiable and that characterizes visually meaningful features according to the human visual system.

4.3 Pictorial style Retrieval and Classification

Research on style-based image retrieval and classification has only started recently. In early work, style-based image retrieval and classification systems mainly depended on brush strokes [45, 57]. Both Melzer et al. [45] and Sablatnig et al. [57] explore the problem of classifying portrait miniatures. Melzer et al. develop two brush detectors, a model based and a neural network approach. Sablatnig et al. build a hierarchical classifier to identify artists, which uses a brush stroke model and consisted of three classification stages, color classification, shape classification and stroke classification.

Recent style-based image retrieval and classification systems have characterized more various low-level features using statistical modelings and machine learning techniques. In addition, the range of applications has extended. Classification between individual artists is a classic example [34, 35, 41, 67], and classifications between photographs and paintings [6, 37, 38] and between photographs and computer graphics images [17] have been explored as well. Leykin and Cutzu [6, 37, 38] distinguish paintings from photographs with a high success rate of over 90%. They use color and texture features and pursue a classifier in an image content-independent manner. Their features include the ratio of intensity edges to color edges, the number of unique colors, the pixel saturation histogram, and the responses of the Gabor filter across multi-orientations and scales. Farid and Lyu [17] build a linear predictor for

the magnitude of the coefficients of the wavelet subbands of natural images based on joint statistics and use it to distinguish between images with hidden messages, computer graphic images, and rebroadcast images. Their linear predictor uses a linear combination of the magnitude of the coefficients of the wavelet subbands of an image. Li et al. [41] compare painting styles of Chinese ink painting artists. They use multi-resolution segmentation to characterize strokes and washes. They link local features with depicted objects using a hidden Markov model.

For European paintings, style-based image retrieval and classification systems have extracted low-level features such as luminance and color distribution [30], frequency properties of spatial blocks [33,34] and even both color and frequency features [44] to build statistical models. These systems use supervised learning to build classifiers for individual artists or art movements. Keren [34] has proposed a classifier of paintings based on local features derived from the discrete cosine transform (DCT) coefficients, which resulted in an 86% success rate on classifying paintings by Rembrandt, Van Gogh, Picasso, Magritte, and Dali. Lombardi et al. [44] use the set of colors in a painting and the frequency distribution of those colors. Their classifier produces an 81 percent success rate. Icoglu et al. [30] have built a classifier that identifies paintings' art movements, such as classicism, impressionism, and cubism. They use a six dimensional feature set and explored three different classification approaches, a Bayesian classifier, a k-NN classifier, and support vector machines. Their features include the percentage of dark colors, color range, skewness of grey scale, number of local and global maxima in the luminance histogram, and so on. Their k-NN classifier shows a high accuracy of over 90%.

Most of these techniques use statistical models and show convincing results. However either they do not provide intuitive characterization [17] or their statistical features are not modifiable and transferable [6, 17, 33, 34, 37, 38, 44].

4.4 Style Transfer

Statistical models are used not only for the measurement of difference or similarity but also for recognition [10, 64] and synthesis. Dror et al. [10] classify surface reflectance properties such as metal, plastic, or paper, based on statistical regularities. They develop image-based reflectance estimation techniques using the pixel intensity histogram and histograms of the coefficients of a bandpass filter pyramid. Torralba and Oliva [64] explore the second-order statistics of natural images and their relevance to image categories. Low-level features are used for the categorization without grouping or segmentation.

Synthesis is another application of statistical models. In particular, parametric synthesis leads to the characterization of textures [25, 36, 50]. Lewis [36] develops a spectral interpretation of textures for digital painting. Heeger and Bergen [25] introduce a method for synthesizing images, matching pixel histograms and the histograms of filter responses at multi-scales and orientations. Portilla and Simoncelli [50] extend this synthesis method to capture correlations between subbands for structured textures.

In recent years, transfer research has focused on methods which copy local textures between images [12, 14, 15, 26] and between strokes [22, 27, 32]. Freeman et al. [22] present one of the first systems for synthesizing line drawings based on example lines. This algorithm fits each input line as a linear combination of several training lines. Kalnins et al. [32] describe stroke-based rendering algorithms. Their system renders silhouettes, synthesizes stroke detail by example, simulates natural media, and hatches with dynamic behavior. These systems inspired the example-based curve synthesis of Hertzmann et al. [27]. Hertzmann et al. also learn line styles from input curves instead of input images.

Much research has explored example-based transfer between images. Efros and Leung [14] show that a simple recursive method of non-parametric sampling can produce new textures similar to examples. Efros and Freeman [15] generate an image by stitching small patches of existing images. Hertzmann et al. [26] learn mapping func-

tions between a pair of images in terms of local filters and apply analogous mapping functions to a target image. Drori et al. [12] introduce a style synthesis technique using example images. They decompose images into overlapping tiles, synthesize each block, and compose the local fragments back together.

Although these studies produce more convincing textures, they use a non-parametric characterization instead of providing statistical models. They cannot be used for analysis nor can they capture large-scale effects. In this thesis, we present a statistical model of pictorial styles. Since the statistical model has explicit parameters, our technique leads to intuitive manipulation and style interpolation.

Chapter 5

Image Features

In this work, we analyze and transfer pictorial styles using quantitative features. We hypothesize that the frequency content of paintings and photographs are related to intuitive notions of style. We study the characteristics of pictures that can be captured by the frequency content across scales, spatial locations, and orientations. In addition, we include features of the low and high residuals of the steerable pyramid. This chapter describes the image features that we choose and qualitatively discusses what they characterize.

We build a statistical model of characteristics of pictorial style, which represent stylistic differences and similarities. We use the steerable pyramid to extract these features. In particular, we focus on three types of features: multi-scale amplitude, the spatial variation of the average amplitude, and the variation of the amplitude across orientations. In this thesis, we concentrate on the stylistic aspects of the luminance channel.

5.1 Multi-scale Amplitude

The use of small-scale texture, sharp lines, and strong contrast between large regions are relevant to pictorial styles. Figure 5-1 shows such examples. Beckmann's painting has a number of sharp lines, and Signac's painting consists of point-shaped brush strokes. These lines and point-shaped brush strokes are the fine-scale or high-

Figure 5-1: The painting by Signac has lots of small dots, whereas those by Beckmann and Chagall do not have a dominant shape or size of brush strokes. Instead, Beckmann's and Chagall's paintings feature visually clear placement of objects.

frequency content. Unlike fine-scale details, the low-frequency description is appropriate for the large-scale contrast and variation of objects and colors of paintings such as in Chagall's work on the right side of Fig. 5-1.

We analyze these multi-scale aspects of pictorial styles in terms of the average amplitude of the steerable pyramid coefficient across scales. We use the following parameters for our analysis of the multi-scale aspects of pictorial styles:

$$\mu_s = \overline{|B_s|}/4^s \ , \tag{5.1}$$
$$\text{where } s = \text{scale.}$$

B_s is the subband of the sth scale of the steerable pyramid. $|B_s|$ stands for the amplitude of the subband coefficient. We denote $\overline{|B_s|}$ as the average of the amplitude. μ_s is the normalized average amplitude of each subband B_s for scales 1 to 5. In this thesis, scale 1 corresponds to $2^8(= 256)$ cycles per image, since the minimum dimension of each image is 512 pixels. Since each subband has a spatial-frequency bandwidth of one octave, an increase by one in the number of the scale means the frequency divides by two. According to the power law of natural images that the average energy falls as "$\frac{1}{f^2}$", so there is equal energy at all octaves. A normalization by 4^s is necessary because the downsampling of subbands is implemented by cropping the discrete fourier transform.

For each image, we build a steerable pyramid with five scales and four orientations. Figure 5-2 shows the representative plots of μ_s (Eq.5.2). The average curves in Fig. 5-2(a) show that the average amplitude curves of different artists have distinct patterns. Moreover, the curves of paintings in Fig. 5-2(b) reveal the regularities shared by the paintings of each artist. The paintings by Chagall and Beckmann, who use strong contrast between large regions, exhibit high average amplitude in low frequencies, whereas Signac's and Monet's paintings have low average amplitude in low frequencies due to their low contrast between large regions. The paintings by Beckmann and Signac have high average amplitude in high frequencies due to their fine-scale features,

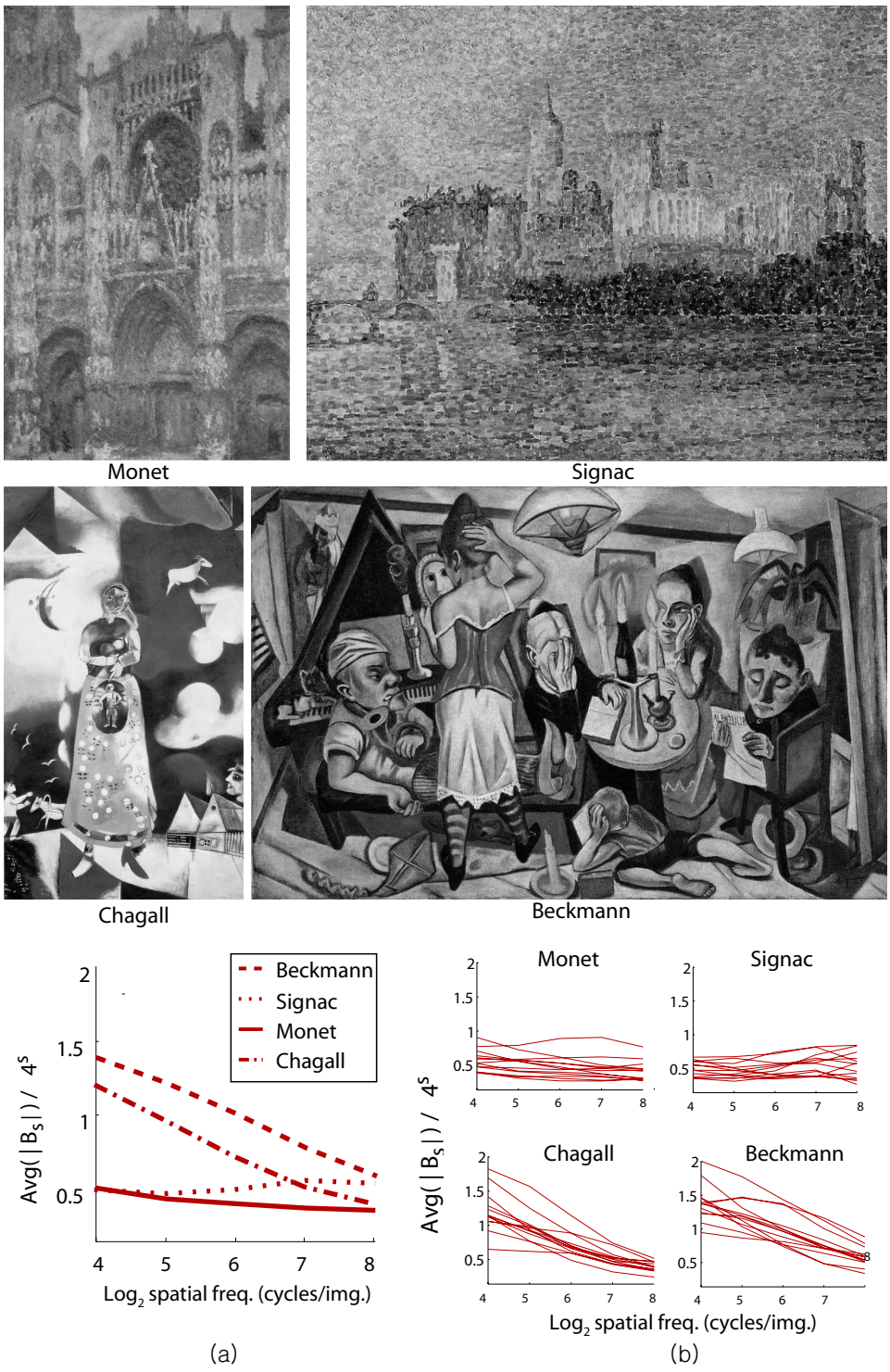


Figure 5-2: (a) The mean of the average amplitude curves of paintings of each artist. The average amplitude curves of different artists have distinct patterns. (b) The average amplitude curves of paintings of each artist. The curves of paintings demonstrate the regularities shared among paintings of each artist.

such as Signac's small dots and Beckmann's sharp lines.

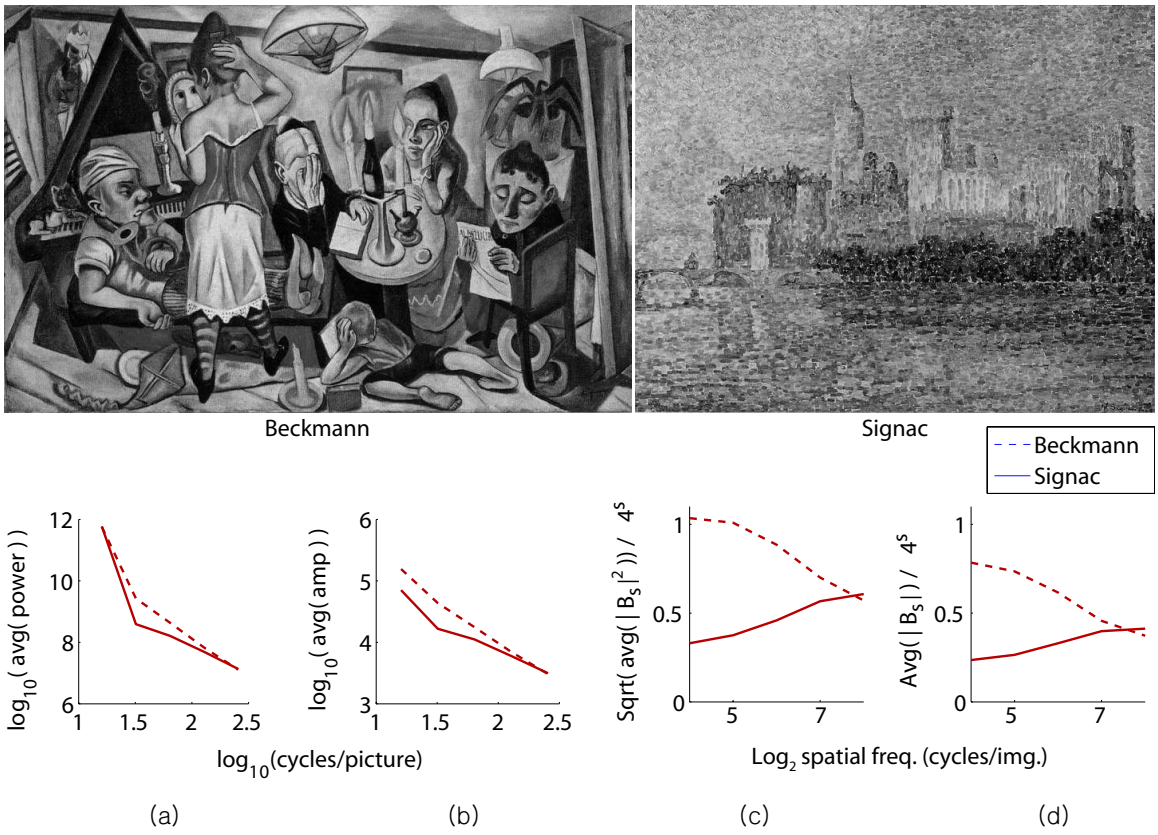


Figure 5-3: Average curves of two paintings of different pictorial styles. (a) The average power spectrum curves in the log-log domain. (b) The average amplitude curves of spectrum in the log-log domain. (c) The scaled square root of the average power curves of the steerable pyramid coefficients in the linear-log domain. (d) The scaled average amplitude curves of the steerable pyramid coefficients in the linear-log domain.

Figure 5-3 shows the average power spectrum curves, the average amplitude curves of spectrum, the scaled square root of the average power curves of the steerable pyramid coefficients, and the scaled average amplitude curves of the steerable pyramid coefficients. Comparison between the average power spectrum in the log-log domain (Fig. 5-3(a)) and the average amplitude of the spectrum in the log-log domain (Fig. 5-3(b)) shows that the average amplitude of the spectrum captures differences between Beckmann's and Signac's paintings more than the average power does. Accordingly, we use the scaled average amplitude of steerable pyramid coefficients (d) as our fea-

tures. In fact, in the steerable pyramids, both of the scaled average amplitude of steerable pyramid coefficients (Fig. 5-3(d)) and the scaled square root of the average power of the steerable pyramid coefficients (Fig. 5-3(c)) capture differences between pictorial styles.

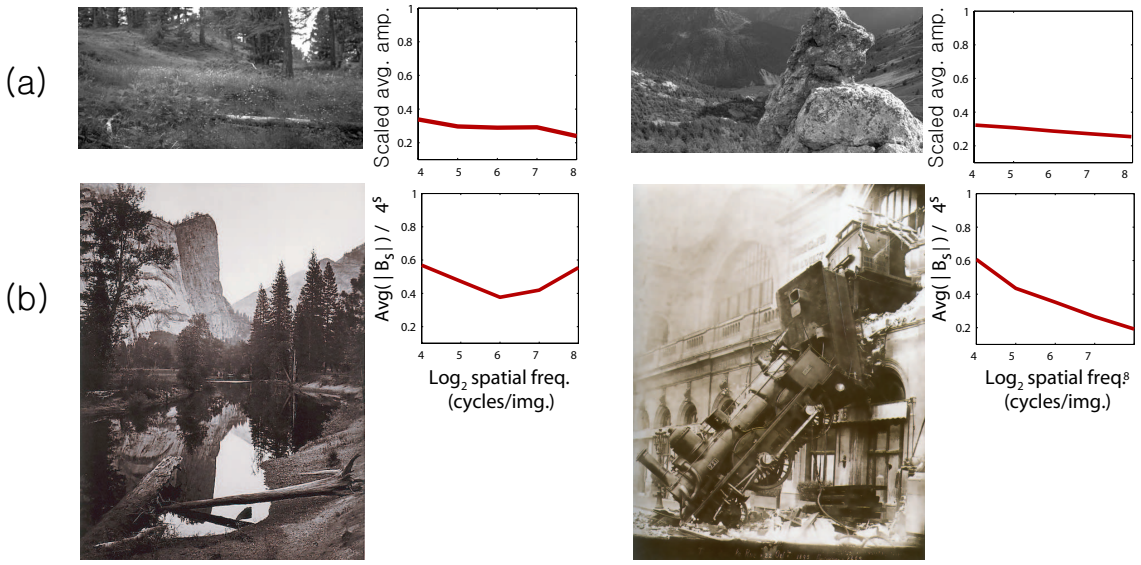


Figure 5-4: The normalized average amplitude across scales: on the left is the low frequency and the right is the high frequency. (a) The average amplitude in casual photographs is uniformly distributed across frequencies. (b) The average amplitude in the artistic photographs shows a unique distribution with a U shape or a high slope.

In addition, the slopes and heights of the average amplitude curves across scales are relevant to the pictorial styles in photographs. Figure 5-4 shows that the differences of frequency content are relevant to the stylistic differences in photographs. The bottom left photograph with the large-scale tonal range from rich black to whitest white and with delicate illustration of fine leaves and trees shows a U-shaped curve of the average amplitude across scales. The high slope of the curve of the bottom right photograph represents the pictorial style of the photograph, which has a blurry classic-movie look. In contrast to these designed photographs, flat-looking casual photographs in Fig. 5-4(a) show uniform distribution of the average amplitude.

As we observed that the curve of the average amplitude across scales characterizes

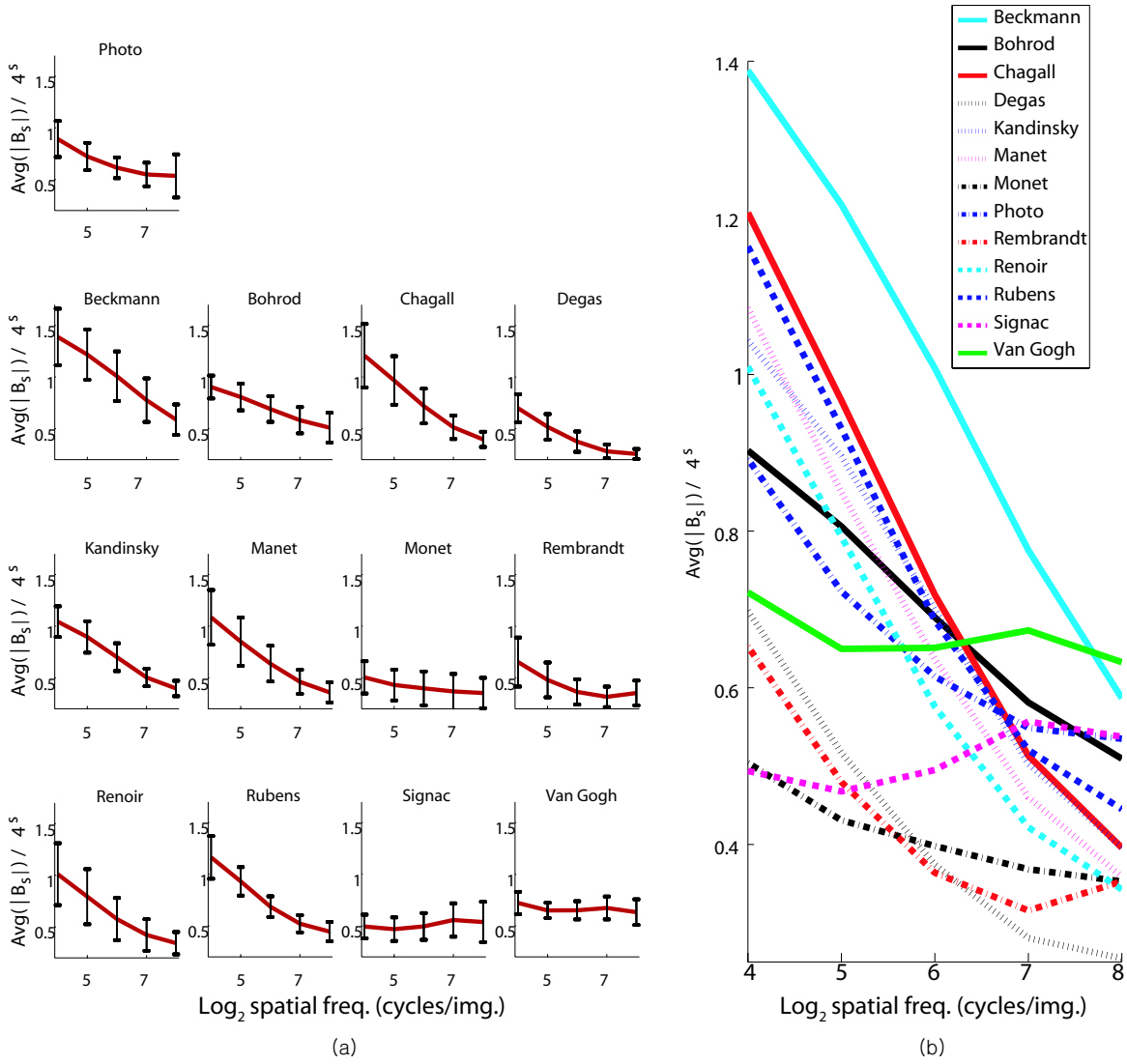


Figure 5-5: (a) The mean of the average amplitude curves of paintings. The standard deviation of each artist shows that the average amplitude curve captures the regularities among paintings of each artist. (b) The mean curves of artists show that the average amplitude characterizes the differences between different artists. Although, in general, each artist has a distinct slope and height, some artists show similar trends.

certain aspects of styles, we explore how the stylistic differences among artists are reflected in the average amplitude. We take the mean of μ_s of eight paintings of each artist and show the mean curves in Figure 5-5. Figure 5-5(a) displays the standard deviations in each scale. The plots show that the average amplitude curve captures the regularities shared by paintings of each artist. Figure 5-5(b) shows that the average amplitude curves characterize the differences among different artists as well. We will discuss the correlation between pictorial styles and the average amplitude curves in the rest of this section.

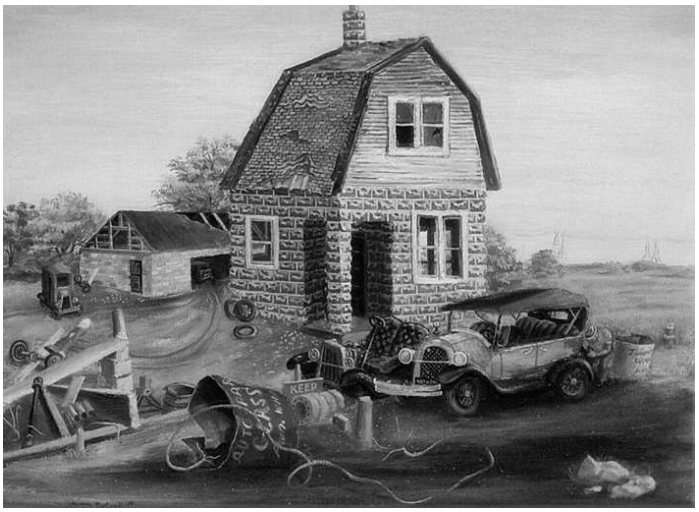
In Figure 5-5(b), Beckmann's curve is well separated from curves of other artists. Beckmann's paintings have high averages in all scales. The high amplitudes result from the sharp edges and lines present in Beckmann's paintings as well as from the strong contrasts between large regions. In comparison, we can observe Monet's and Signac's low and relatively flat curves in their average amplitude plots. Their low amplitudes in low frequencies result from the absence of coarse-grain contrast. Also, note that Signac's curve features high energy in the high frequencies due to his point-shaped brush strokes.

The average heights of the average amplitude of Degas' and Rembrandt's paintings are as low as those of Monet and Signac. However, Degas' and Rembrandt's paintings do not have flat average curves; instead they have average amplitude curves with slopes. Degas and Rembrandt do not use as small brush strokes as Monet and Signac but paint with more contrast between large regions. In fact, Monet, Signac and Degas are Impressionists, and Rembrandt's paintings belong to the Baroque style. Composition and subject matter might separate Degas from Monet and Siganc and group Degas and Rembrandt. Many of the paintings by Degas and Rembrandt have people in them, while the paintings in our data set, by Signac and Monet are landscapes. Figure 5-6 shows their differences.

The paintings by Chagall, Manet, Renoir, Kandinsky, Rubens, Bohrod and Van Gogh, and photographs have about the same height of the average amplitude curves. However, their slopes differ. Van Gogh has a flat average curve. Since Van Gogh uses visible shapes and sizes of brush strokes, the amplitudes in the high frequencies

Figure 5-6: Signac's and Monet's paintings have low contrast between large regions, and Degas and Rembrandt have higher large-scale contrast.

Van Gogh



Bohrod

Chagall

A photo by Sander Studio

Figure 5-7: The differences in the slopes of the average amplitude curves among artists in Fig. 5-5 are relevant to their uses of brush strokes, the oriented structure like lines and edges, and the contrast between large regions.

are higher than those of the other artists. Photos and paintings by Bohrod do not have as steep slopes in their average amplitude curves as in other artists, such as Chagall, Manet, Renoir, Kandinsky and Rubens who have similar heights of curves. Photos and the paintings by Bohrod have large-scale contrasts that are higher than Van Gogh, but lower than Chagall, Manet, Renoir, Kandinsky and Rubens. This leads to their moderate heights in the low frequencies. Regarding high frequency features, photos and paintings of Bohrod do not have visible marks of brush strokes like those by Van Gogh, yet they have sharper edges and lines than other paintings. Those sharp edges and lines cause heights in high frequencies to be higher than those of Chagall, Manet, Renoir, Kandinsky and Rubens, but less than those by Van Gogh. (See Figure 5-7.)

So far, we have discussed the relevance of the height and slopes of the curve of the average amplitude across scales and pictorial styles. The average amplitude can represent visual differences of the use of brush strokes, oriented structures such as lines and edges, and large-scale contrast.

5.2 Non-stationarity

In the previous section, we have studied the multi-scale aspects of pictorial styles using the steerable pyramid. We have shown that the average amplitude across scales is relevant to the regularities and differences of pictorial styles. However, the average amplitude curve captures global statistics of pictorial styles and disregards the variation of the statistics due to spatial location. In this section, we discuss the aspects of pictorial styles that are related to the spatial variations of multi-scale content.

Some paintings have the same degree of detail over the entire image, while in others the degree of detail varies spatially. Figure 5-8 shows example paintings that have different degrees of spatial distribution of details. Rembrandt's painting has some detailed regions and other regions that are less detailed. In comparison, Beckmann's painting has a uniform distribution of details over the image. Images have different degrees of the spatial variation of detail for many different reasons. The distribution

Figure 5-8: Rembrandt's *The Blinding of Samson* and Beckmann's *Family Picture*. Rembrandt's painting has some regions with lots of details and other regions with less detail. In comparison, Beckmann's painting has uniform distribution of details over the image.

of the local details is used to draw attention. Also, the variation across the spatial locations may be intended for the illustration of the realistic perspective. An effective use of the spatial variation can be used to compensate for the flatness of two dimensional canvas.

Although our model cannot characterize the purposes of the spatial variations in pictorial styles, it can parameterize the visual effects according to the spatial variations. To this end, we first examine the local spectra signatures. The spatial variations of images correspond to the variation of statistical properties of images, which leads to the variation of local spectra signatures. In image statistics, this variation is called non-stationarity. Figure 5-9 shows the correspondence between the non-stationarity of images and their local spectra signatures. A large spatial variation in an image leads to a large variation of local spectra signatures.

Since we observed that the local spectra are relevant to the non-stationarity of styles, we explore how the stylistic differences between artists are reflected in the spatial variation of average amplitude. We take a $k \times k$ size window $W_{s,i}$ of B_s around each pixel x_i and measure the average amplitude within each subspace $\overline{|W_{s,i}|}$ for all i. We compute the standard deviation of $\overline{|W_{s,i}|}$ over all i and normalize it by the average amplitude of B_s to prevent high average amplitudes from leading to high

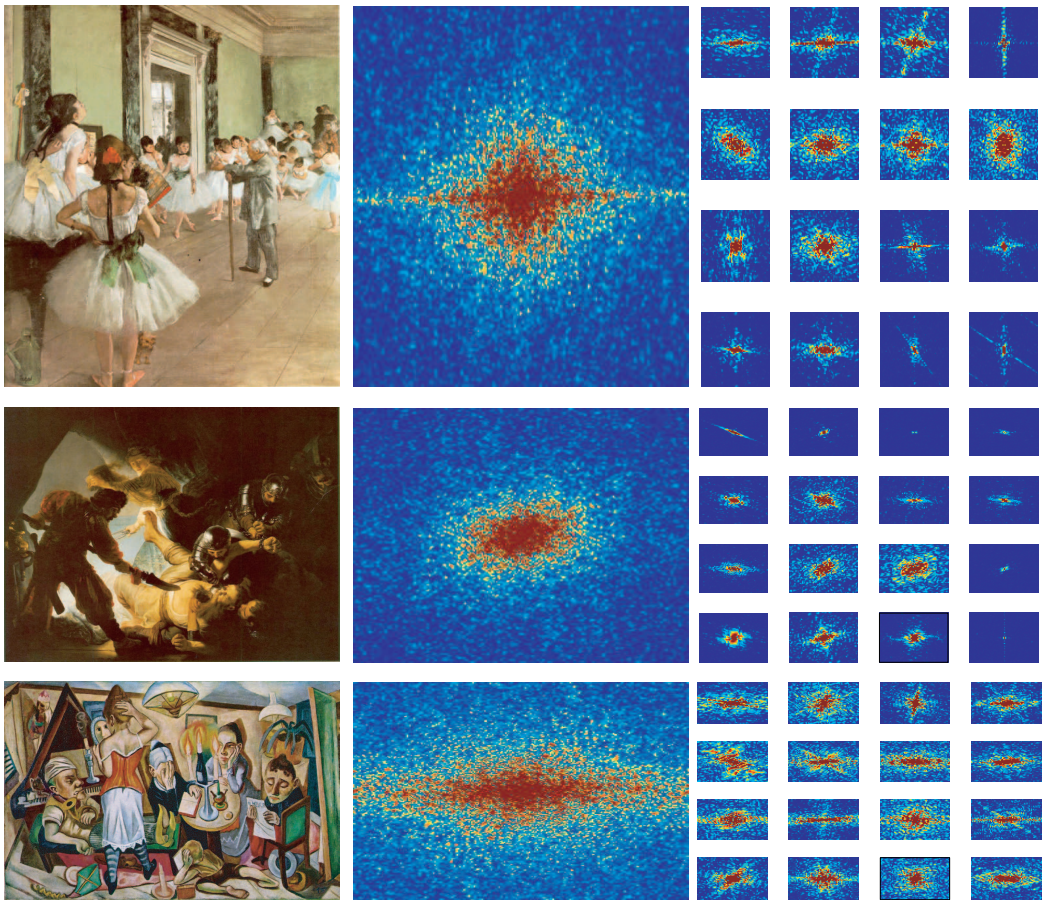


Figure 5-9: The local spectra signatures of images are relevant to the pictorial styles. A large spatial variation in an image leads to a large variation of local spectra signatures.

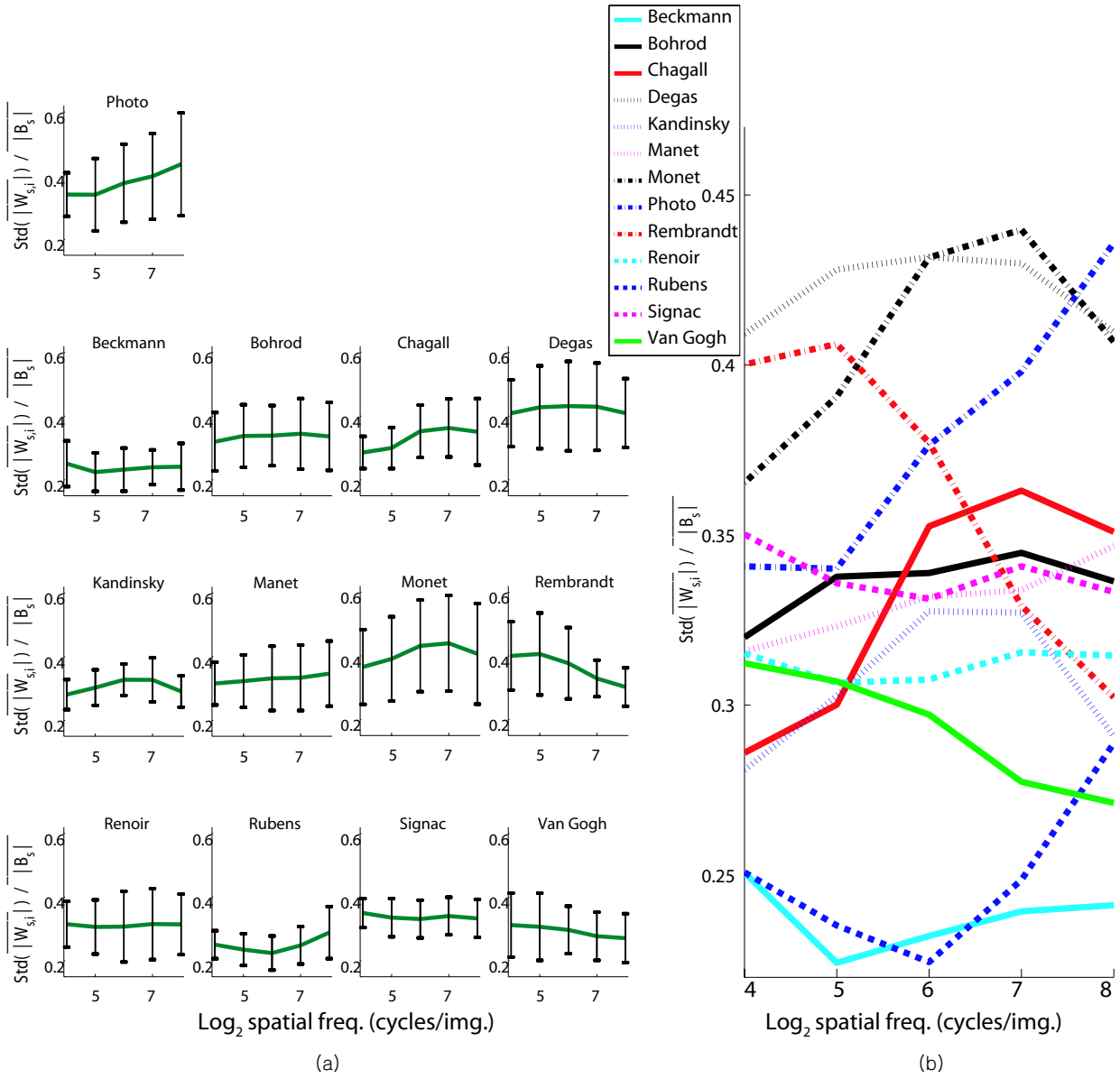


Figure 5-10: (a) The mean of paintings' σ_s, the standard deviation of the average amplitudes of the windows normalized by the average amplitude in the whole image. The standard deviation of each artist shows that our non-stationary parameters capture the regularities among paintings of each artist. (b) The mean curves of artists have various slopes and heights correlated to the differences between pictorial styles of artists. A high curve represents a large variation of detail in the whole image, such as in Degas's paintings. In contrast, some artists like Beckmann have low curves due to the stationarity present in their pictorial styles.

standard deviation regardless of the actual variation.

$$\sigma_s = std(\overline{|W_{s,i}|})/\overline{|B_s|}\ , \tag{5.2}$$
$$i = \text{spatial location and } s = \text{scale.}$$

We use $k = \frac{128}{2^{s-1}} + 1$ at each scale s, from 1 to 5. A smaller k would divide an image into many small windows and it is likely that the variations of these windows would not represent signatures of pictorial style. Instead, the variations of such windows might characterize random features, which are not relevant to pictorial style. Furthermore, a larger k would result in large windows, each of which would not capture local properties of frequency content. Since each window covers a large region, we might lose features relevant to pictorial styles by taking an average amplitude of each window.

We plot the mean of σ_s of paintings of each artist in Figure 5-10. Figure 5-10(a) displays the standard deviations in each scale. The plots show that the standard deviation of the average amplitude of windows captures the regularities shared among paintings of each artist. The standard deviation of our non-stationarity feature over paintings is higher than the standard deviation of our multi-scale feature over paintings. For some artists such as Degas and Monet, the variation is quite high. Since these artists are Impressionists who use colors more than lines or brightness to express their subject, our study using the luminance channel might miss important spatial color contrast.

Figure 5-10(b) shows that the non-stationarity σ_s characterizes the differences among different artists. For example, Rembrandt shows a very distinctive curve with its sharply descending slope. Rembrandt used localized spatial composition. However, Rembrandt's lines and brush strokes are not visible. The response of the steerable pyramid to them is not strong in any region. In addition, Rembrandt's dark background might generate noisy responses. These non-localized high frequencies lead to low standard deviations.

Monet Degas

Rubens Beckmann

Figure 5-11: Monet and Degas show a high spatial variation of detail due to the presence of the sky and the background. In comparison, Rubens' and Beckmann's paintings have uniform a distribution of details.

Over all scales, Degas and Monet exhibit high spatial variation. In fact, the cause of the high spatial variation is different for the two artists. Monet's paintings are mostly landscapes with sky which leads to high spatial variation. Degas' paintings are not usually landscapes but have people in them. However, how Degas places objects and people locally generates the high spatial variation. In comparison, we can notice Beckmann's and Rubens' low curves. Beckmann's and Rubens' paintings have details and objects evenly distributed over whole regions in the images, which leads to low variation in the spatial domain. Figure 5-11 illustrates these differences and similarities.

We have discussed the spatial variation aspects of pictorial styles. The spatial variation of the average amplitude is used to measure the non-stationarity. The spatial variations are related to the subject matter, composition, use of background, and so on.

5.3 Anisotropy

In the previous sections, we have shown that the average amplitude across scales is relevant to pictorial styles. The high amplitude in the coarse scales corresponds to the contrast between large regions over an image, and the amplitude in the fine scales is generated by the high frequency details, such as fine-scale brush strokes and oriented structures like lines and edges. However, the analysis of multi-scale aspects of pictorial styles does not distinguish the source of the high amplitude in a certain scale. For example, the high amplitude in the high frequencies does not distinguish whether it is from non-oriented structures, such as textures or point-shaped brush strokes, or from oriented structures such as lines and edges. Therefore, a measure of the orientedness is necessary to distinguish between non-oriented structures and oriented structures. We explore the anisotropy aspects of pictorial styles using the standard deviation of the amplitude across orientations.

The orientedness or anisotropy represents whether or not the features in the images are structured and aligned along particular directions. Lines and edges are

Figure 5-12: Signac's *The Papal Palace, Avignon* and Beckmann's *Family Picture*.

examples of oriented structures. Figure 5-12 demonstrates the difference between pictorial styles with non-oriented features and those with oriented features. Signac's *The Papal Palace, Avignon* does not have explicit lines, although we can see the boundaries of the building. In contrast, Beckmann's *Family Picture* has real line marks to represent line structures. The steerable pyramid we use is appropriate to analyze such oriented structures, since we can locate the clear lines where the oriented filters respond strongly. The analysis of the response of the filters reveals the degree of the orientedness in an image.

We use the standard deviation of the amplitude across orientations to analyze the characteristics of the "orientedness." The parameters used in the analysis are the following:

$$\rho_s = \overline{std(|B_{s,k}(i)|)}/\overline{|B_s|} \ , \tag{5.3}$$

$$s = \text{scale},\ i = \text{pixel location, and } k = \text{orientation.}$$

We have the oriented slice $B_{s,k}$ of each subband B_s in the steerable pyramid for the kth orientation. At each point x_i, we compute $std(|B_{s,k}(i)|)$, the standard deviation of $|B_{s,k}(i)|$ across four orientations. Then, we take the average of the standard deviations over the whole image. We normalize this orientation feature by the average amplitude

of B_s to prevent high average amplitudes from leading to high standard deviation. We compute ρ_s at scale 1 to 5.

The high standard deviation across orientations is related to the existence of edges and lines, but our anisotropy feature is not sufficient to distinguish between features such as texture and edges and lines. Edges possess additional properties not captured by anisotropy such as coherence across scale and spatial coherence along the edge.

As we observed that the orientedness is relevant to certain aspects of pictorial styles in Figure 5-12, we explore how the stylistic differences between artists are reflected in the standard deviation of amplitude across orientations. We take the mean of ρ_s (Eq. 5.4) of paintings of each artist and show the mean curves in Figure 5-13. Figure 5-13(a) displays the standard deviation of ρ_s among paintings of each artist. They show that the average amplitude curve captures the regularities shared by paintings of each artist. In addition, Figure 5-13(b) shows that the mean of the standard deviations across orientations characterizes the differences between different artists as well. In the rest of this section, we discuss the correlation between pictorial styles and the mean of the standard deviations across orientations.

Beckmann and Kandinsky, whose paintings contain sharp edges and lines, have high curves and the highest peaks in the $4th$ scale. In contrast, the paintings by Monet and Signac show the highest peaks in the $2nd$ scale and have low curves. This is related to their pictorial styles not having lines and edges at fine scales. In addition, in Van Gogh's curve in Figure 5-13, the height of the curve is not as high as the height of Beckmann's curve, but the curve has a very clear peak in the fourth scale. This is relevant to the fact that Van Gogh uses very strong lines with a specific width. In fact, Van Gogh's spin-whirl brush strokes generate quite thick marks, which are captured by a certain scale.

The anisotropy features distinguish high-frequency content from oriented structures and those from non-oriented textures or brush strokes. Signac has shown as much high-frequency contrast as Beckmann in Figure 5-5, but in Figure 5-13, Signac does not have as high values in the high frequencies as Beckmann does. Using anisotropy features, we can distinguish the differences of pictorial styles, which could

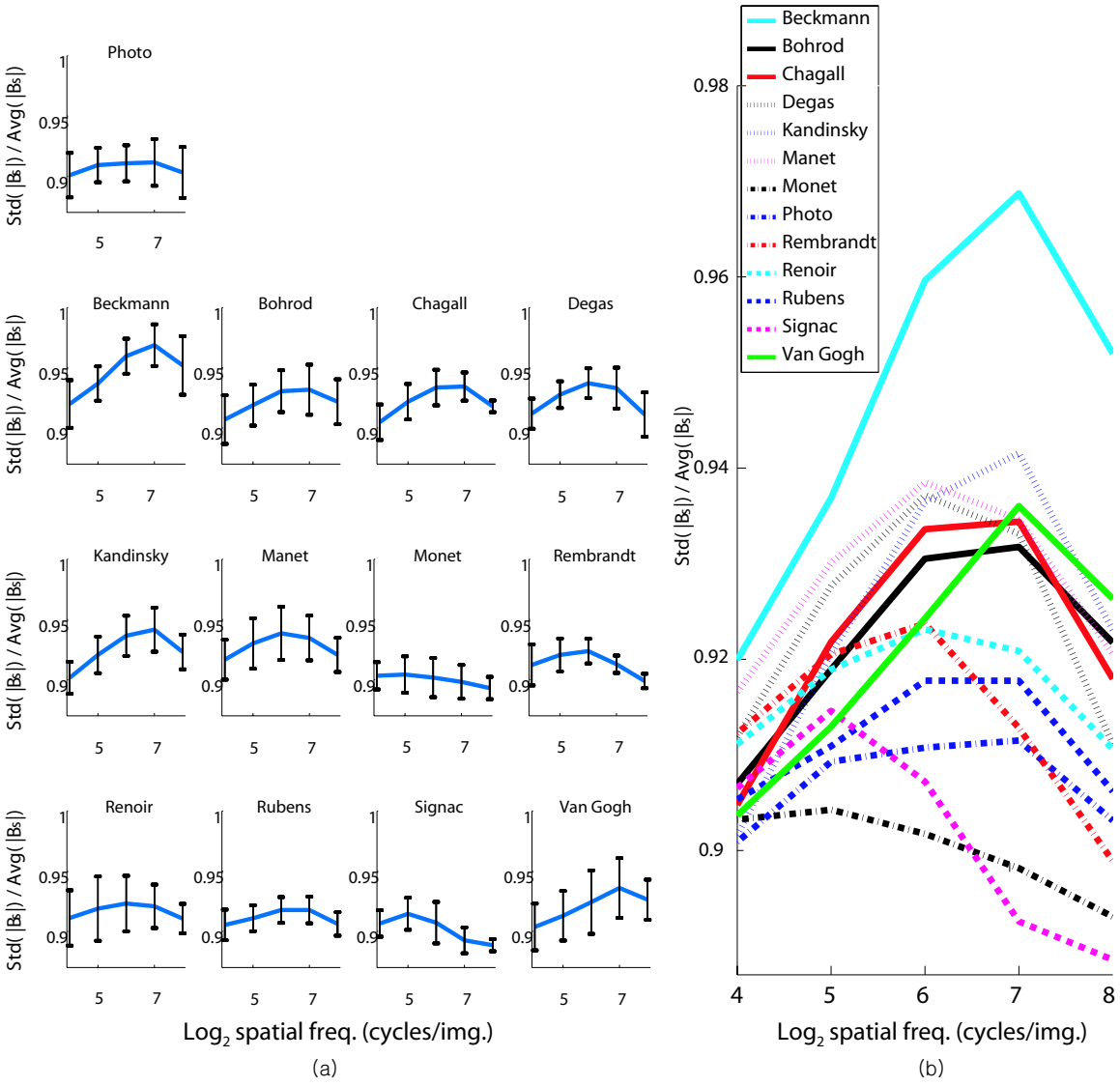


Figure 5-13: (a) The mean of the average amplitude curves of paintings. The standard deviation of each artist shows that the average amplitude curve captures the regularities among paintings of each artist. (b) The mean curves of artists show that the average amplitude characterized the differences between different artists.

not be done by other features.

5.4 High and Low Residual features

In addition to our three main types of features, we include features of low and high residuals of the steerable pyramid for completeness. We take the average amplitude of the high residual and take a mean and a standard deviation of the low residual. We do not analyze these three features in depth as we do for our other fifteen features. In particular, the low residual is relevant to the pixel histogram, which we do not analyze in this thesis, and the effect of the high residual is rather redundant as it has similar values as the high frequency subband.

5.5 Summary

In this chapter, we have described three main types of features that characterize pictorial styles using the steerable pyramid as well as residual features. We have shown that the multi-scale amplitude is relevant to the use of brush strokes, oriented structures such as lines and edges, and large-scale contrast. Furthermore, the spatial variation of the average amplitude can measure non-stationary aspects of pictorial styles, that is, the degree of spatial distribution of details. In addition, we have shown that the variation of the amplitude across orientations captures the differences between oriented structures such as lines and edges and non-oriented structure such as textures and brush strokes.

Chapter 6

Style Analysis

In the previous chapter, we have discussed how our image features are relevant to pictorial styles. In this chapter, we will analyze these features using machine learning techniques to further examine the relevance of our features to regularities and differences of pictorial styles and to discuss correlations and dependencies among the features.

Our eighteen features capture the multi-scale, non-stationarity, and anisotropy aspects of images. However, it is still difficult to visualize and understand eighteen-dimensional data. Moreover, the features are neither independent nor uncorrelated, and the examination of paintings along an individual feature does not reveal the underlying regularities. Machine learning techniques can be used to reduce the number of dimensions for the purpose of analysis and more importantly to extract the meaningful projection of features, i.e. to reveal correlations and dependencies.

We employ two kinds of machine learning techniques: Principal Component Analysis (PCA), which is an unsupervised learning technique, and Linear Discriminant Analysis (LDA), which is a supervised learning technique. Both of them reduce the dimensionality of features by projecting the original feature vectors onto a subspace. The features we selected are the amplitude distribution across scales, the amplitude distribution across the spatial domain, and the standard deviation of amplitude across orientations.

We use PCA to visualize how the features are relevant to natural style classifica-

tion. It is natural to classify style by artist. However, it should be understood that each artist does not always retain a uniform style. Using PCA not only helps validate the selected features, but also demonstrates the directions in the feature space along which data points are well-spread. In addition, PCA decorrelates features so that we can observe certain aspects that we would not be able to analyze. Therefore we examine the principal directions to understand the significance of each feature and the correlation between features.

Unlike PCA, LDA is a supervised learning technique. That is, we assign the artist labels to the paintings of two different artists, then LDA separates paintings of different artists while clustering paintings of the same artist. To this end, the LDA maximizes the ratio of the distance between groups to the distance within group. LDA is more useful than PCA to extract the projection of features that is actually relevant to the classification and labeling.

6.1 Principal Component Analysis (PCA)

In this section, we validate the relevance of our features to regularities and differences of pictorial styles. Using PCA, we transform the feature space into a principal component space according to the variance of features. The direction with the largest variance among paintings and photos in the feature space is assigned to be the first principal direction in the transformed space. First, we examine each type of feature, multi-scale, non-stationarity, and anisotropy with PCA. In this way, we can separately analyze the relevance of individual features of each type. Once we understand each type of feature, we analyze all features together.

We do not label the feature vectors by artists. That PCA spreads paintings of different styles and that PCA does not spread paintings of similar styles validate the relevance of the features to the pictorial styles. However, that the feature vectors of paintings by one artist are not clustered does not necessarily mean that the features are not appropriate for capturing regularities. Artists explore different pictorial styles, and those images with different styles tend to be more spread than clustered. In

addition, that the paintings by different artists are clustered together does not mean that the features are not characterizing the stylistic differences. Along one type of notion of pictorial styles, different artists might share the same statistical properties.

6.1.1 Multi-scale Analysis

Among the three types of features, multi-scale, non-stationarity, and anisotropy, the multi-scale features separate the images by artists the most. In particular, the multi-scale features are meaningful to capture the differences and similarities in paintings. Artists have many degrees of freedom to change the multi-scale content. They can explore the sizes of brush strokes, contrast between large regions and so on.

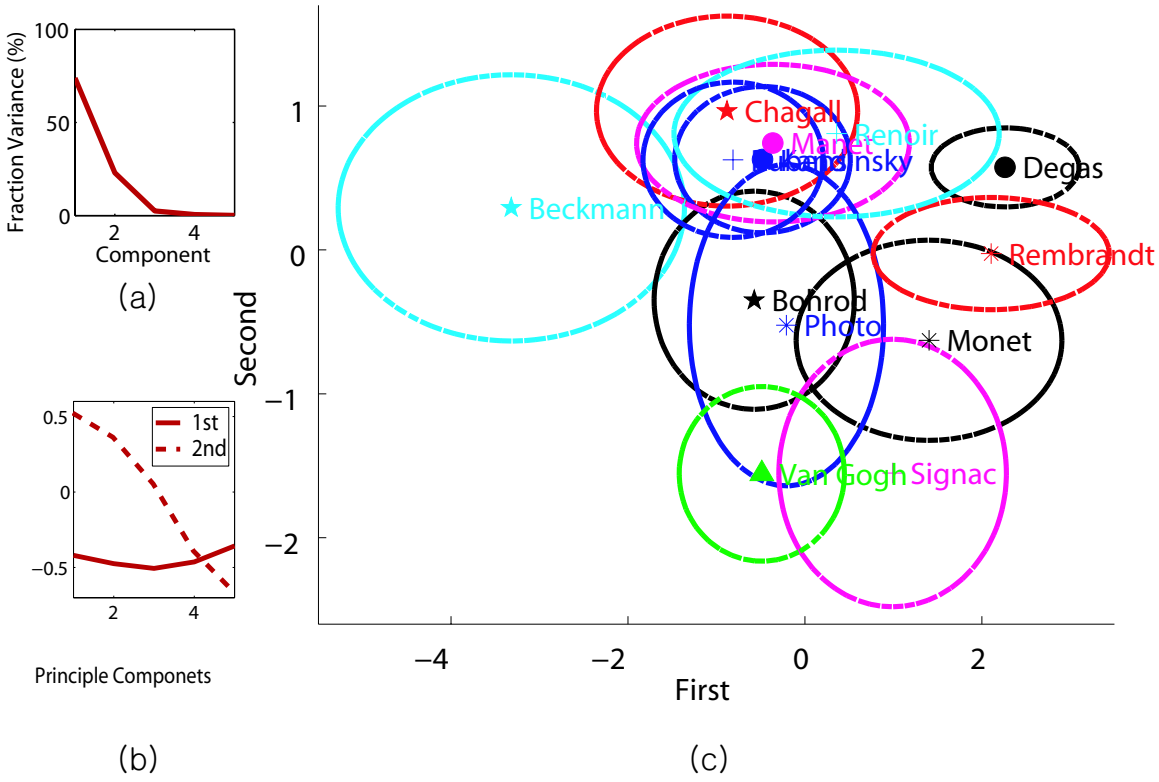


Figure 6-1: The first and second principal components regarding the multi-scale features. (a) The fraction variance of each component. (b) The first and second principal components. (c) The mean and the standard deviation of paintings by each artist. The ellipses represent the standard deviation of the components among the paintings by each artist.

Figure 6-1 shows how paintings and photographs are spread along the first and second principal components regarding the multi-scale features. The first principal component is the height of the average amplitude curve. This represents how much details are present generally over all scales. The second component is its slope. As the fraction variances of components show in Figure 6-1 (a), the other three components do not spread paintings as the first two components do.

Paintings by artists approximately fall into six groups in Figure 6-1 (c). The first group of Beckmann, the second group of Chagall, Renoir, Manet, Rubes and Kandinsky, the third group of Bohrod and Photo, the fourth group of Degas and Rembrandt, the fifth group of Monet and Signac and the sixth group of Van Gogh.

The first component places the second (Chagall, Renoir, Manet, Rubens and Kandinsky), third (Bohrod and Photo), sixth groups (Van Gogh) in the similar position, and the fourth (Degas and Rembrandt) and fifth (Monet and Signac) groups together.

The paintings by Degas, Rembrandt, Monet and Signac tend to have less luminance contrast. Instead, some of them have more contrast in colors, which we do not study in this thesis. In comparison, the paintings by Chagall, Renoir, Manet, Rubens, Kandinsky, Bohrod, and Photographs have more noticeable luminance contrast. Figure 6-2 shows example paintings by artists in the relative positions according to the first and second principal components. We can observe that different degrees of luminance contrast spans the space along two principal components.

In addition, notice that the first component separates Beckmann's paintings from the paintings of different artists apart. Beckmann's paintings have sharp edges and lines as well as strong contrast between large regions. These high contrast features distinguish Beckmann's paintings from the other paintings.

The paintings of Monet, Signac and Van Gogh have low large-scale contrast compared to the paintings by Chagall, Manet and Renoir. This aspect of pictorial styles is reflected in their positions of the second principal component.

Another interesting observation is that the paintings of Bohrod and photographs are plotted in the similar position along the first and second principal components. In

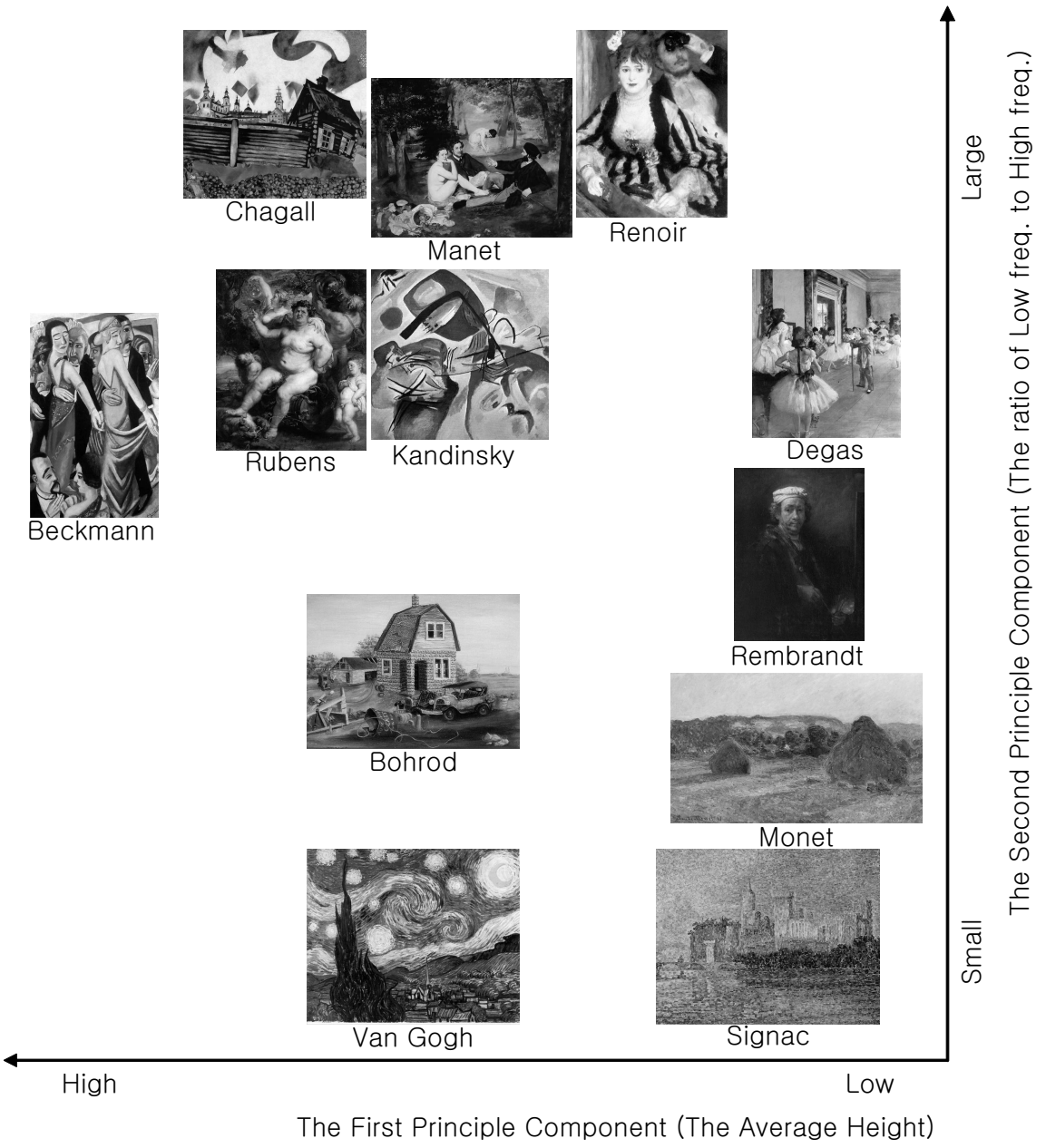


Figure 6-2: Example paintings by artists in the relative positions according to the first and second PCA components regarding the multi-scale features

terms of content, Bohrod's work is classified as Surrealism. However, regarding our multi-scale aspects of pictorial styles, the photographic looking of Bohrod's paintings place them with photographs.

As we observed, the paintings are spread by the principal components of the multi-scale features in a way that they are separated according to the differences and similarities in terms of visual and intuitive pictorial styles. Moreover, the degree of intersection between different artists reflects how these two artists share some aspects of pictorial styles. For example, the large intersection between Renoir and Degas is relevant because work by both of them is categorized as Impressionism. Another intersection between two other Impressionists, Signac and Monet is such an example as well. The co-location of Chagall's and Kandinsky's paintings is somehow natural, considering that both of them are called Expressionists. However, notice that Manet and Renoir are positioned very close to Chagall and Kandinsky, although Manet and Renoir are Impressionists. As we have mentioned, our features are not designed to draw clear boundaries between conventional art movements. Instead, we capture the visual differences such as smooth brush strokes, strong luminance contrast and contrast between large regions. Therefore, it is natural that Manet and Renoir, who used smooth brush strokes and high luminance contrast, are clustered with Chagall and Kandinsky.

6.1.2 Non-stationarity Analysis

Non-stationarity is high in an image where some regions have more details than the rest of the image. The non-stationarity features are more appropriate to analyze photographs where depth of field plays an important role. Depth of field results in blurred out-of-focus regions and sharpened focused regions. In paintings, the depth of field effect is not as strongly employed by all artists. In addition, the variation of details depends on the subject matter. Images with the sky tend to have large variation of details because of the absence of detail in the sky. In addition to the sky, images with a large background are likely to have large variation. The non-stationarity features depend on whether the image has large background and whether

the image is a photograph.

Figure 6-3 shows the first and second components of PCA for the standard deviation across spatial locations. The first principal component is the height of paintings' standard deviation across spatial locations. A high value in this axis implies the image has high variation of details at all scales. The second principal component is the slope of paintings' standard deviation across spatial locations, that is, whether the image changes the degree of detail variation across scales. As the fraction variances of components show in Figure 6-3 (b), the other three components do not spread paintings as the first two components do.

In Figure 6-3, paintings and photographs are separated into four groups. The first group of Beckmann and Rubens, the second group of Van Gogh, Signac, Kandinsky, Renoir, Bohrod, Manet, and Chagall, the third group of Rembrandt, the fourth group of Degas, Monet and Photo. The space of the first and second component does not generate clear boundaries between groups. However, the fact that the groups are not clearly separated does not mean these non-stationarity features does not capture pictorial styles. We discuss the relevance of the features to the differences and similarities of pictorial styles in the rest of this section.

The first component plots Beckmann and Rubens in similar positions, and Rembrandt, Degas, Monet and Photo close to one another. The low first components of Beckmann and Rubens are relevant to their uniform distribution of details in their paintings, whereas the dynamic compositions of Rembrandt, Degas, Monet and Photo result in the high standard deviation across spatial locations. We do not intend to select specific contents or composition of these paintings and photos. However, in our data set, most of Beckmann's and Rubens's paintings do not have background, but are full of people or objects. In comparison, many of the paintings of Rembrandt, Degas and Monet and photographs in our data set have the sky or background. (See Figure 6-4.) In particular, Rembrandt places objects locally, and a large part of his paintings is background. This means that low-frequency content is not uniform across spatial locations. However, his high-frequency content is not captured due to the absence of sharp lines and visible brush strokes. This leads to non-localized high-frequency

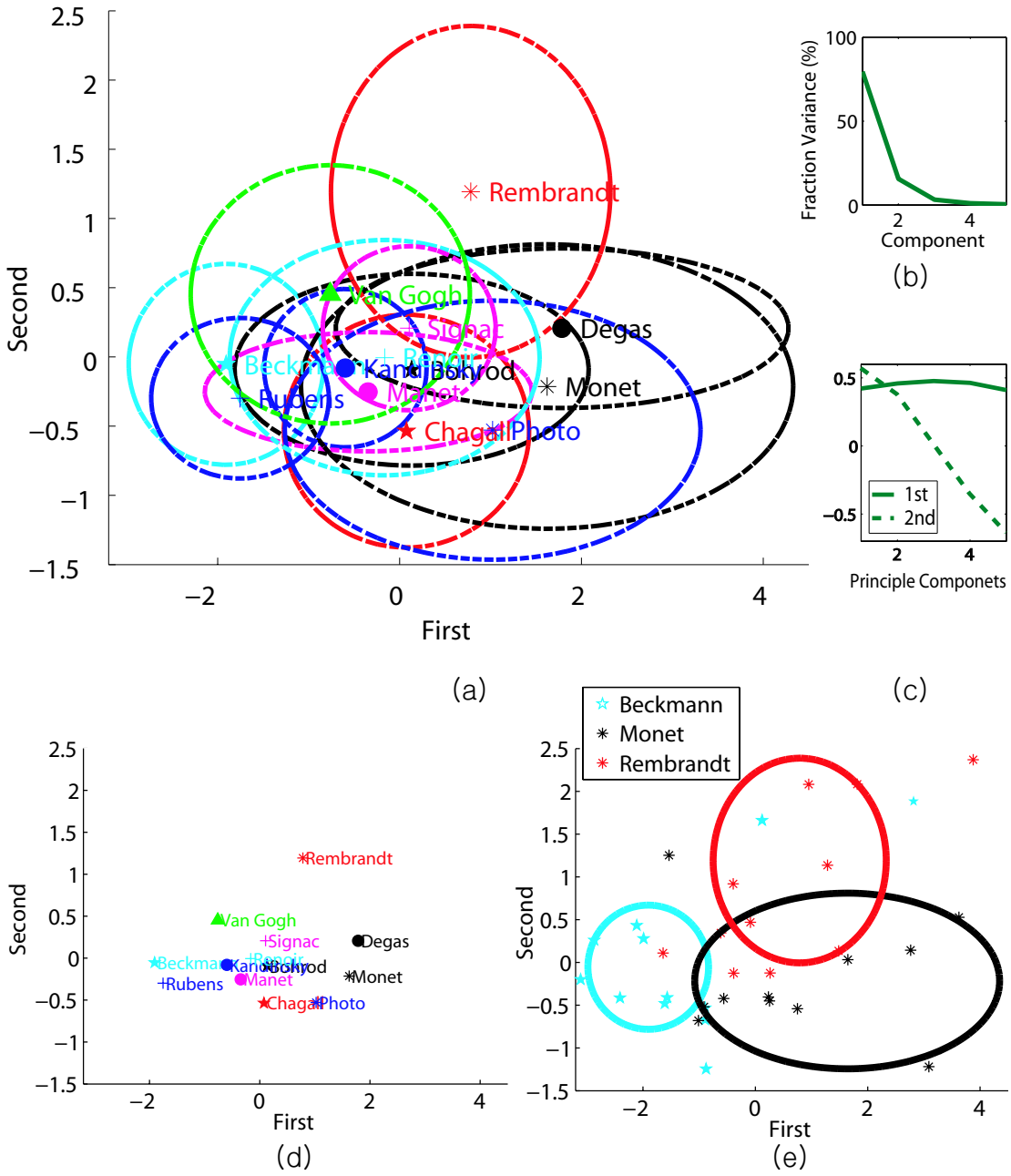


Figure 6-3: The first and second components of PCA for the standard deviation across spatial locations. (a) The mean and the standard deviation of paintings by each artist. The ellipses represent the standard deviation of the components among the paintings by each artist.(b) The fraction variance of each component. (c) The first and second principal components. (e) Marks represent paintings.

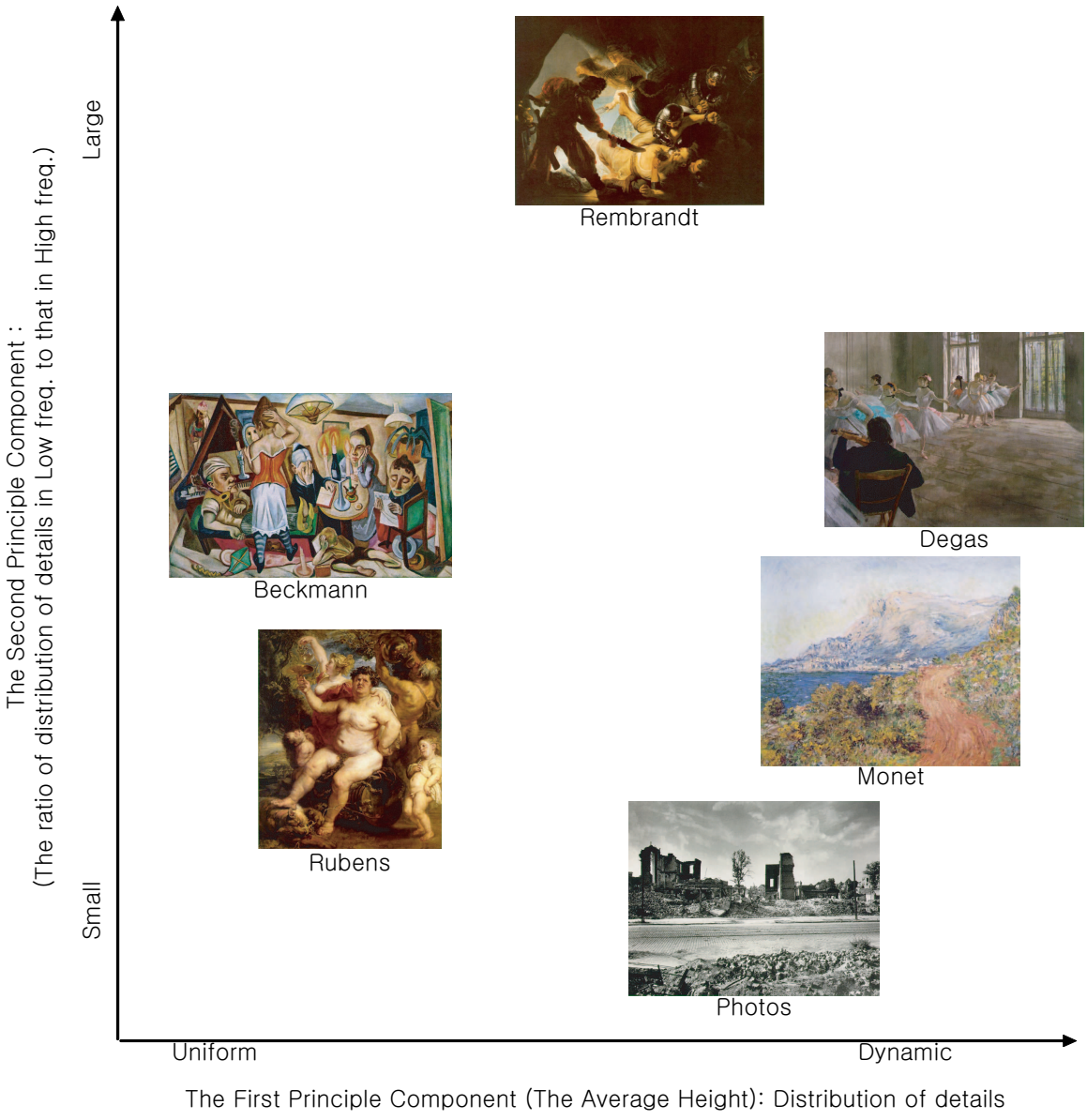


Figure 6-4: Example paintings by artists of the variation of non-stationarity

content.

Although non-stationarity features does not provide clear separation of paintings by artists, these features certainly capture visual differences and similarities between pictorial styles.

6.1.3 Anisotropy Analysis

In the principal component analysis for anisotropy features, the height of paintings' standard deviation across orientations is the first principal component and the slope is the second component. (See Figure 6-5.) As the fraction variances of components show in Figure 6-5 (b), the other three components do not spread paintings as the first two components do.

The anisotropy features separate images into five groups in Figure 6-5. The first group of Beckmann, the second group of Van Gogh, Chagall, Bohrod, Manet and Kandinsky, the third group of Rubens and Photo, the fourth group of Degas and Renoir, the fifth group of Rembrandt, Monet and Signac.

The first group of Beckmann's paintings is separated well again from the paintings of different artists: Beckmann's paintings have high standard deviation across orientations in all scales. The high standard deviations result from the oriented structures, such as sharp edges and lines. However, notice that paintings by Beckmann are separated into two groups in Fig. 6-5(d). In fact, Beckmann's paintings have two distinctive pictorial styles. Figure 6-6(a) are some example paintings separated apart and (b) shows the example paintings close to Signac's data points.

Among our features, anisotropy features separate paintings in the most intuitive way. The human visual system is sensitive to the oriented images features such as lines and edges. That the paintings by Degas and Renoir are grouped, the paintings by Rubens and photographs are grouped, and the paintings by Manet, Chagall, Bohrod, Kandinsky, and Van Gosh are grouped seem very natural.

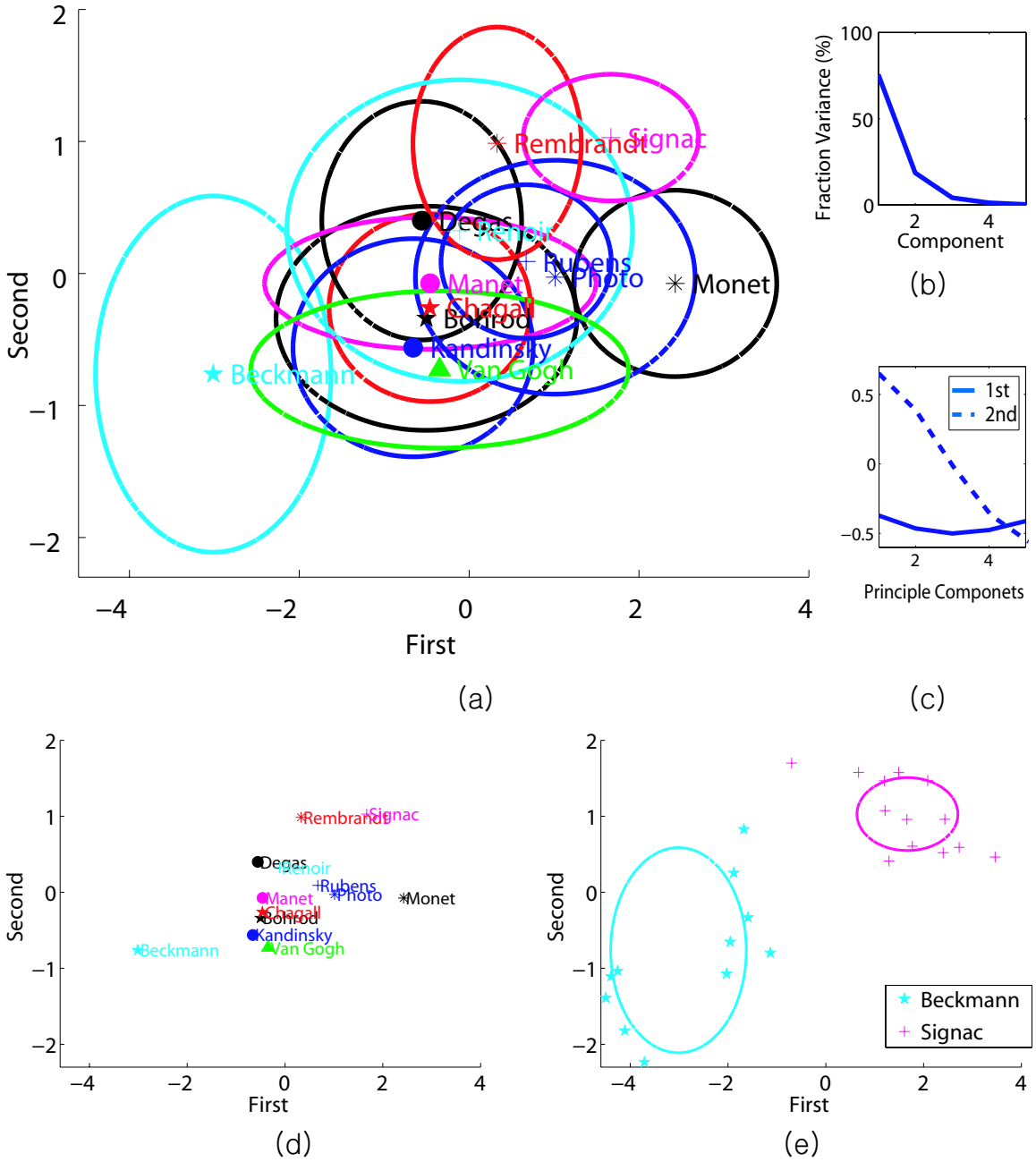


Figure 6-5: The first and second components of PCA for the standard deviation across spatial locations. (a) The mean and the standard deviation of paintings by each artist. The ellipses represent the standard deviation of the components among the paintings by each artist.(b) The fraction variance of each component. (c) The first and second principal components. (e) Marks represent paintings.

(a)

(b)

Figure 6-6: Beckmann's paintings of different pictorial styles.

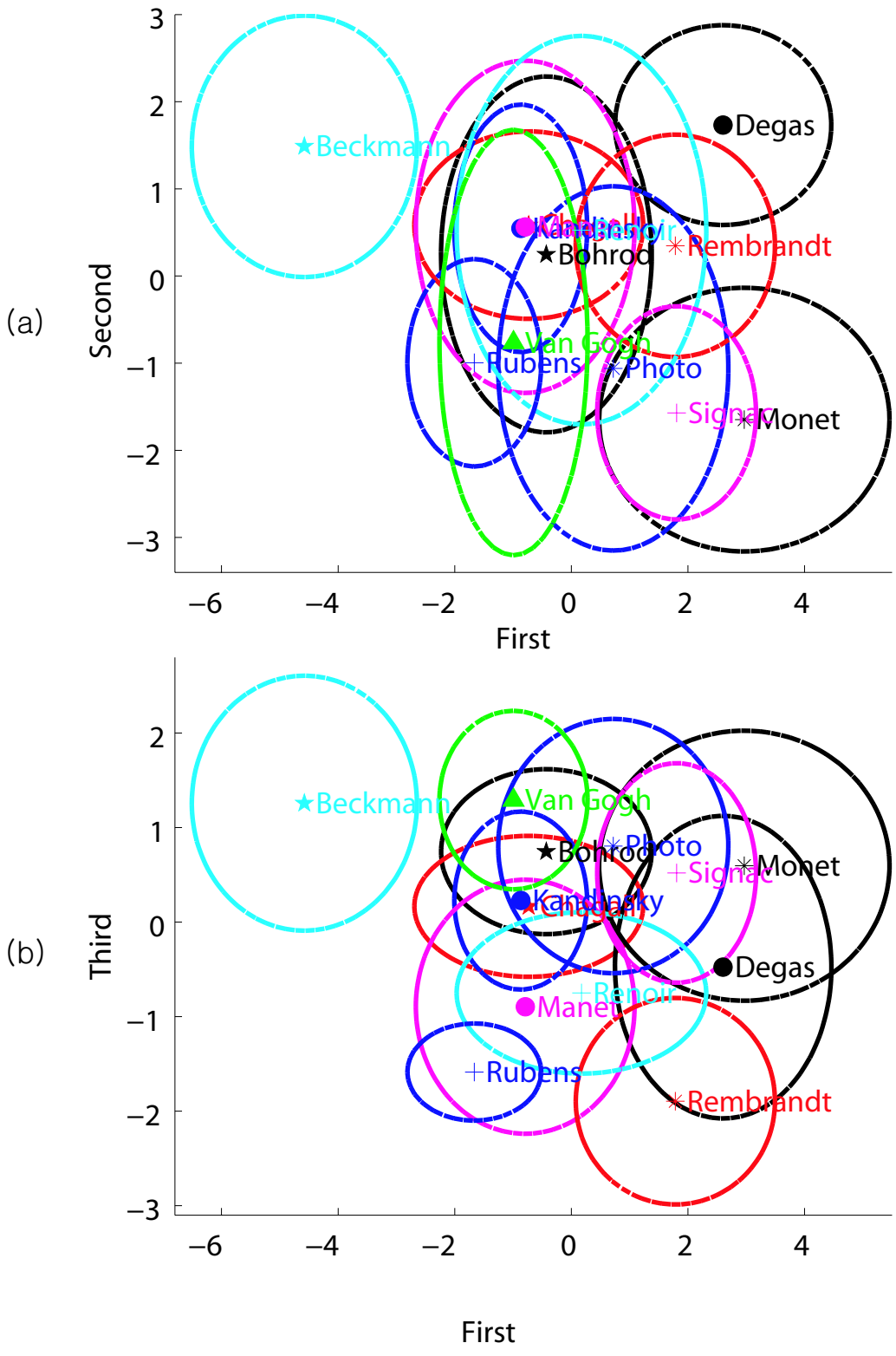


Figure 6-7: (a) The first and second components of PCA. (b) The first and third components of PCA. Oval shapes represent the standard deviation of component values. The ovals show how spread paintings by each are around the average value of the components.

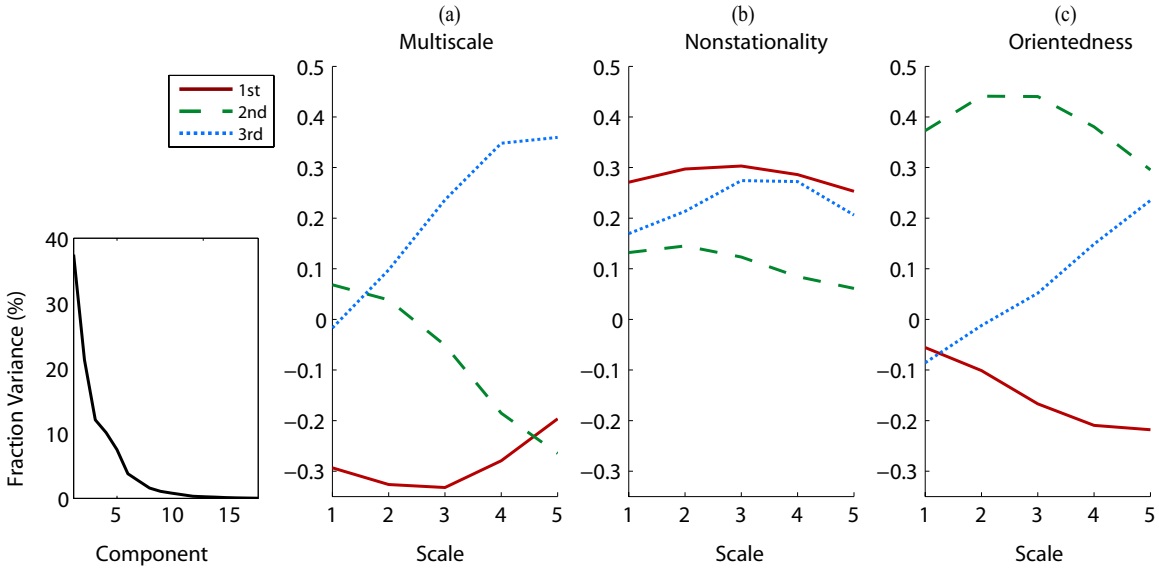


Figure 6-8: (a) The three principal components of PCA for the multi-scale features (b) The three principal components of PCA for the non-stationarity features (c) The three principal components of PCA for the anisotropy features.

6.1.4 Correlation and Dependency Analysis

In the previous sections, we have studied the fifteen features separately. We now put the fifteen features together and add the features of high and low residuals, which are the average amplitude of the high residual and the average and standard deviation of the low residual amplitude of the steerable pyramid. We perform PCA on these eighteen features, as shown in Figure 6-7 and 6-8.

For the average amplitude feature, the first component captures the height of the average amplitude curve and the second and third components capture the slope of the average curve. For non-stationarity, the height determines the first component and most of the second component. The slope determines the third component and a part of the second component. In the anisotropy, the slope and the location of the peak of average of standard deviation across orientations contribute to the first and second components. The third component is about the slope of the average standard deviation across orientations.

Compared to the previous individual analysis of each type of features, the influ-

ences of heights and slopes are changed. In particular, for the anisotropy features, the first component is the slope, whereas it is the height when we analyze only the anisotropy features. The first component for the multi-scale features remains the same as it is the height. There might be correlation between this change and scaling issues. As we normalize features, scaling does not affect the analysis of features of one type. However, each type of feature is expressed with its own "units" and the scale between them is more or less arbitrary. We will explore this scaling issue in future study.

In figure 6-7 (a), the first and second components project paintings in a way that the paintings of each artist are clustered. However, it is possible that the non-separable boundaries result from the fact that each artist does not retain a uniform style and two different artists possibly share certain aspects of styles. In fact, the third component separates some artists clustered along the second component. For example, Manet and Renoir are separated from Kandinsky and Chagall in the third component, while they are clustered with them in the second component. This separation is possible since the non-stationarity in the third component tells the difference between these two groups, Manet and Renoir, and Chagall and Kandinsky.

Principal components project the statistical characteristics of each artists in Fig. 6-7. Artists who have visually different styles are separated in the space of the principal components. For example, Beckmann's paintings are well separated from the other paintings. Signac and Monet's paintings place close to each other. As Chagall and Kandinsky share many aspects of styles, such as color schemes, composition, brush strokes and so forth, Chagall and Kandinsky's paintings are clustered.

Our principal component analysis has validated that our features capture regularities of paintings with similar pictorial styles and the differences of paintings with distinctive pictorial styles. Our features classify paintings not by artists, but by coarse-grain pictorial styles. In addition, photographs are clustered although they are not separated from paintings. This is because some artists like Bohrod follow pictorial styles looking like photographs. In terms of our features, the statistical features of photographs and the paintings of Bohrod are not separable.

	bec.	boh.	cha.	deg.	kan.	man.	mon.	pho.	rem.	ren.	rub.	sig.	van.
bec.	0.0	1.5	2.4	2.5	1.8	2.2	8.8	1.9	4.9	2.3	3.6	28	4.4
boh.	**1.5**	0.0	0.7	1.4	0.8	1.2	1.8	0.8	6.8	1.3	1.9	8.2	**1.2**
cha.	2.4	0.7	0.0	3.2	**0.4**	0.7	2.6	1.1	4.8	1.2	1.5	23	2.4
deg.	2.5	1.4	3.2	0.0	3.4	1.4	4.8	1.7	**1.7**	**0.7**	2.3	9.6	2.5
kan.	1.8	**0.8**	**0.4**	3.4	0.0	2.7	8.6	0.7	7.7	1.7	1.7	*29*	4.2
man.	2.2	1.2	0.7	1.4	2.7	0.0	3.5	0.4	2.2	**0.7**	1.1	5.4	4.1
mon.	8.8	1.8	2.6	4.8	8.6	3.5	0.0	**0.4**	13	1.2	3.7	1.7	5.4
pho.	1.9	0.8	1.1	1.7	0.7	**0.4**	**0.4**	0.0	2.5	0.7	**1.0**	**2.0**	2.5
rem.	4.9	6.8	4.8	1.7	7.7	2.2	*13*	2.5	0.0	2.3	2.6	20	*19*
ren.	2.3	1.3	1.2	**0.7**	1.7	0.7	1.2	0.7	2.3	0.0	1.1	4.1	2.7
run.	3.6	1.9	1.5	2.3	1.7	1.1	3.7	1.0	2.6	1.1	0.0	7.5	12
sig.	*28*	*8.2*	*23*	*9.6*	*29*	*5.4*	1.7	2.0	*20*	*4.1*	7.5	0.0	7.4
van.	4.4	1.2	2.4	2.5	4.2	4.1	5.4	*2.5*	19	2.7	*12*	7.4	0.0

Table 6.1: The measure of class separability

6.2 Linear Discriminant Analysis (LDA)

Unlike PCA, LDA is a supervised learning technique that exploits labeling of the data, by artists in our case. It extracts the direction in the multi-dimensional feature space that best distinguishes the artists. It tells us which features or combination of features are most effective to separate two artists apart. We performed LDA on the same set of eighteen features used for PCA: multi-scale features, non-stationarity features, anisotropy features, and the features for high and low residuals. LDA helps to extract the main axes of variation, which are related specifically to the differences between a pair of artists.

Table 6.2 shows the measure of the class separability between pairs of artists among 12 artists and that between each artist's paintings and photographs. The separability is the ratio of the distance between groups to the distance within group. Small numbers mean the two artists have very similar statistical properties. The smallest number for each artist is marked with a bold font and the largest number is denoted with an italic font. Notice, Chagall and Kandinsky show very small separability, which is relevant to the similarity of their stylistic properties. Signac and Kandinsky mark the largest separability due to their different pictorial styles.

In most cases, the anisotropy features are assigned the largest absolute weights. Large absolute weights mean that the features are effective at discriminating between

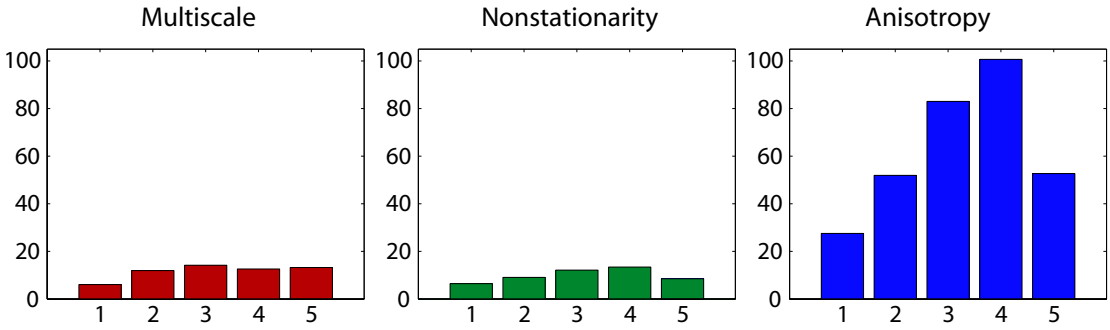


Figure 6-9: Averaged absolute values of LDA weights for the multi-scale, non-stationary and anisotropy features.

the two artists, whereas weights close to the zero mean that two artists are not separable along the axis. Figure 6-9 shows the average absolute weights in our Linear Discriminant Analysis results. We observe that the weights of the anisotropy features are larger than the other features. This means that artists are separable along our anisotropy features consistently. Let us discuss each pair case by case.

The plots in Figure 6-10 and 6-11 show LDA plots for ten pairs of artists. The separability is the ratio of the distance between groups to the distance within group. (a) shows how the paintings of those two artists are separated. (b),(c) and (d) are for LDA weights assigned to the multi-scale, non-stationary and anisotropy features. We display pairs with large separability in Figure 6-10 and those with small separability in Figure 6-11. In both cases, anisotropy features are assigned with the largest weight. In addition, notice that the easily-separable pairs assign much higher weights to the anisotropy features than non-separable pairs. This suggests a strong correlation between pictorial styles and the anisotropy features. In particular, we hypothesize that this features might characterize the scale at which strong edges occur, which might be an important signature that distinguishes paintings by an artist from others. We will further study the relative size of weights between features in future work.

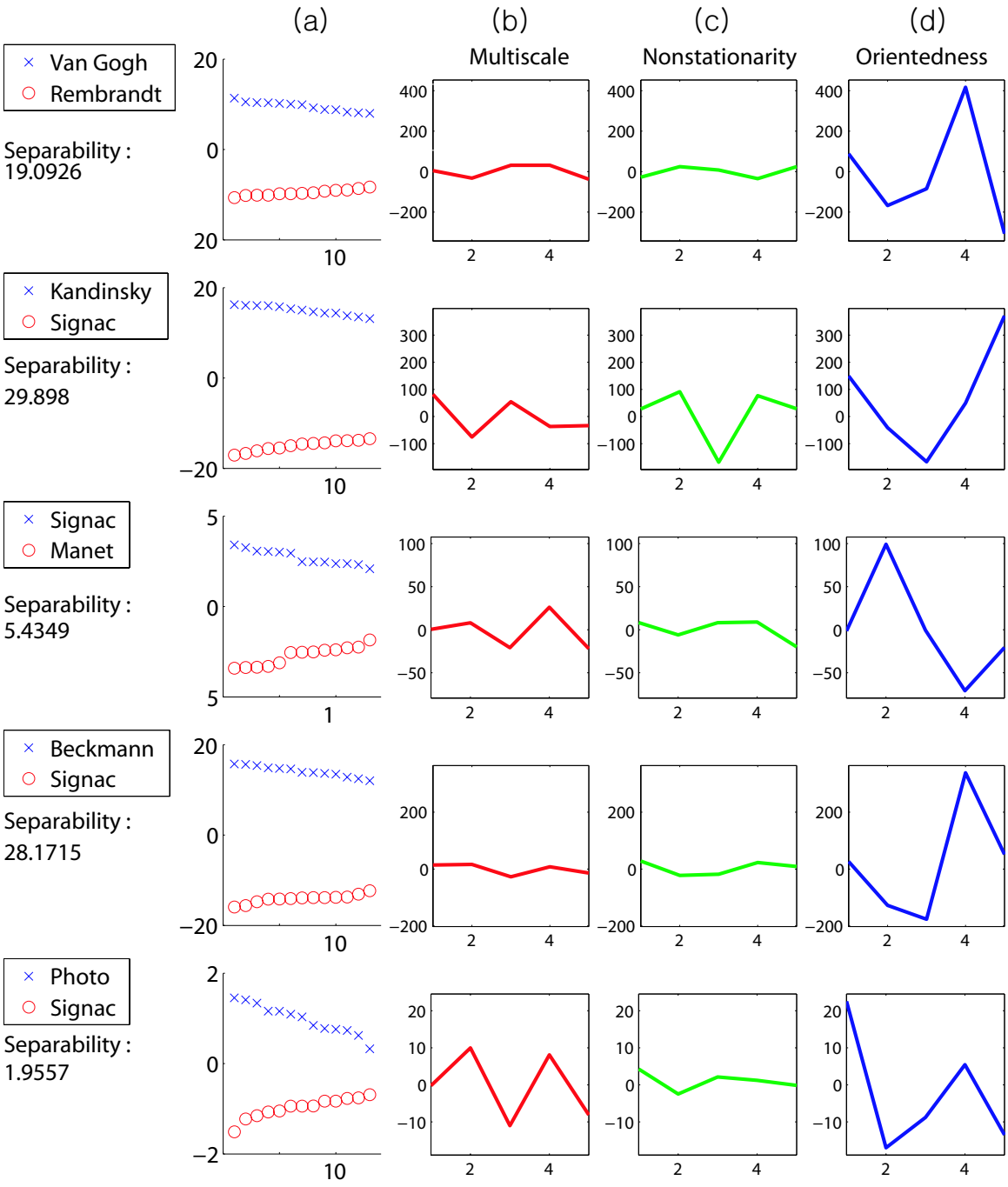


Figure 6-10: LDA plots for five pairs. The separability is the ratio of the distance between groups to the distance within group. (a) shows how the paintings of those two artists are separated. (b),(c) and (d) are for LDA weights assigned to the multi-scale, non-stationary and anisotropy features.

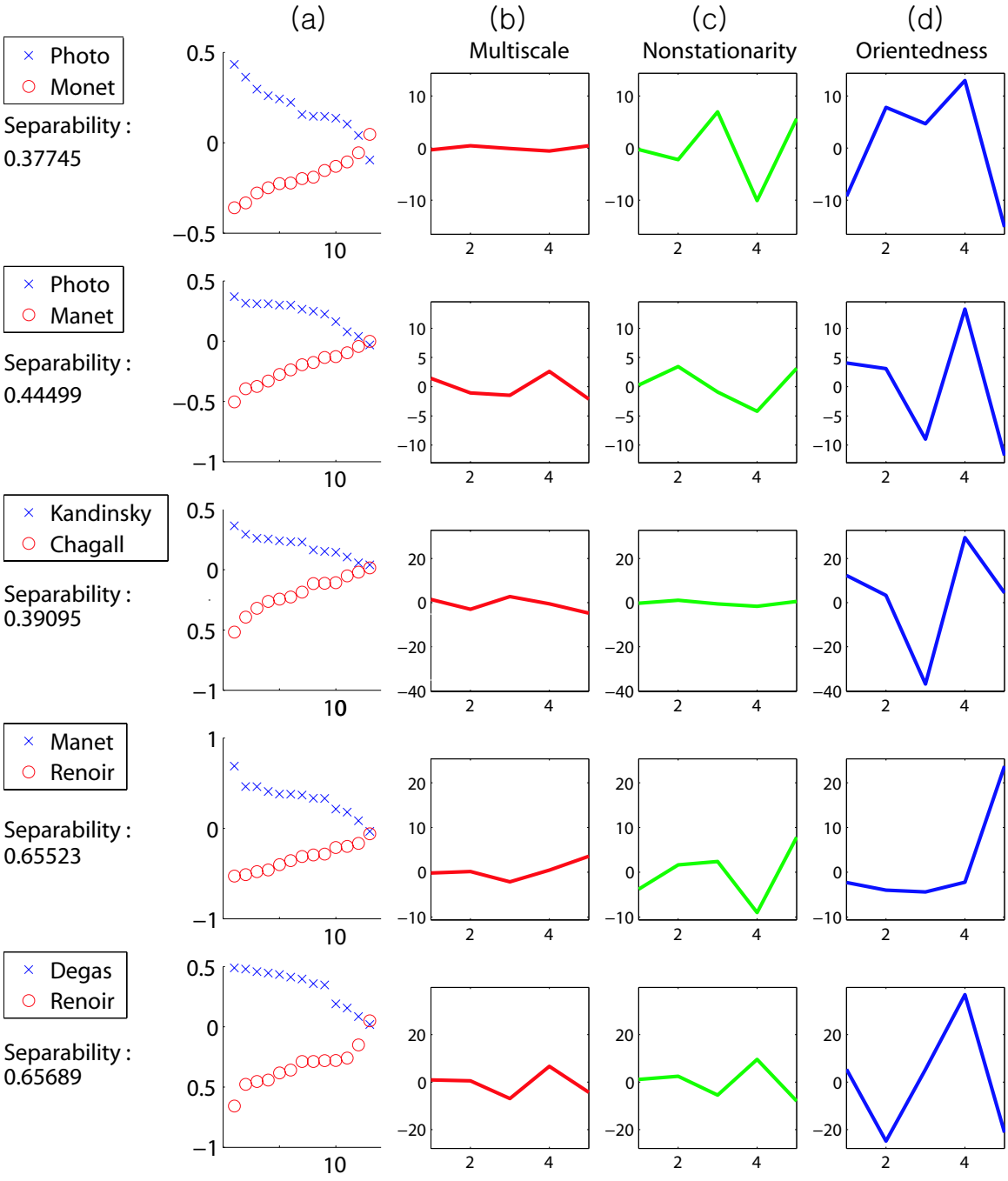


Figure 6-11: LDA plots for five pairs. The separability is the ratio of the distance between groups to the distance within group. (a) shows how the paintings of those two artists are separated. (b),(c) and (d) are for LDA weights assigned to the multi-scale, non-stationary and anisotropy features.

6.3 Summary

In this chapter, we have analyzed our features using machine learning techniques, PCA and LDA. The results of PCA and LDA validate that our features are relevant to regularities and differences of pictorial styles. PCA clusters paintings by one artist and separates paintings by artists with different pictorial styles. LDA measures the class separability between a pair of artists, which is consistent with visual differences between pictorial styles of two artists.

Chapter 7

Style Transfer

In the previous chapter, we studied the relevance of the features to stylistic differences using PCA and LDA. In this chapter, we assess the visual relevance of these statistical features by transferring them between pairs of images. We match eighteen features which capture multi-scale, non-stationary and anisotropy aspects of images.

We believe that transfer is an important tool for the evaluation of statistical image features. While our transfer results are far from perfect from an application standpoint, they allow us to assess the relevance and orthogonality of our feature set, and they suggest where future work is needed.

We compare the results obtained by our whole feature set with those by each individual feature. In addition, we discuss the limitations and advantages of our method as compared to previous methods such as Simoncelli and Portilla's textures analysis and synthesis [50], Heeger and Bergen's histogram matching [25], and Hertzmann et al.'s image analogies [26], although some of them have not been developed for style transfer.

The transfer results without certain features assess the compactness of our features. The difference between results with and without the features validates that the features are necessary. The changes of different features are designed to be independent to one another. We include the transfer between two images that have similar features as a sanity check.

The evaluation of transfer is difficult because subjective judgement is involved.

Therefore, we try to transfer styles between images with similar content. For example, transfer between a landscape image and a portrait might not meet the expectation of viewers.

7.1 Multi-scale Content Transfer

As we have shown in the previous chapters, the average amplitude curve across scales captures multi-scale stylistic aspects of an image. The transfer of the average amplitude changes the degree of large-scale contrast and the strengths of high-frequency content such as textures, lines, and edges.

We transfer the multi-scale average amplitude by multiplying coefficients by the ratio between an input image and a model image at each scale from 1 to 5. This gives the input image a visual look similar to the model image. B'_s is the subband of the s *th* scale of the steerable pyramid of the model image. B_s is the subband of the s *th* scale of the steerable pyramid of the input image. Our transfer generates the following new subband for the input image:

$$B_s^{new} = \frac{\overline{|B'_s|}}{\overline{|B_s|}} B_s, \tag{7.1}$$
$$\text{where } s = \text{scale}.$$

After this multi-scale transfer, we apply pixel histogram matching, as Heeger and Bergen transfer [25] does.

Figure 7-1 shows the transfer result. The scale transfer, Fig. 7-1 (c) and the the whole features transfer (d) achieve the large-scale contrast of Chagall's painting and remove the point-shaped brush strokes of Pointillism, which cannot be done by intensity matching (e). The scale transfer (c) and the Heeger and Bergen transfer (f) create similar results, although the scale transfer (c) transfers only average amplitudes and (f) transfers full histograms [25].

We show the transfer result with the opposite direction in Figure 7-2, which transfers features from the painting by Signac to one by Chagall. In Figure 7-2, the scale

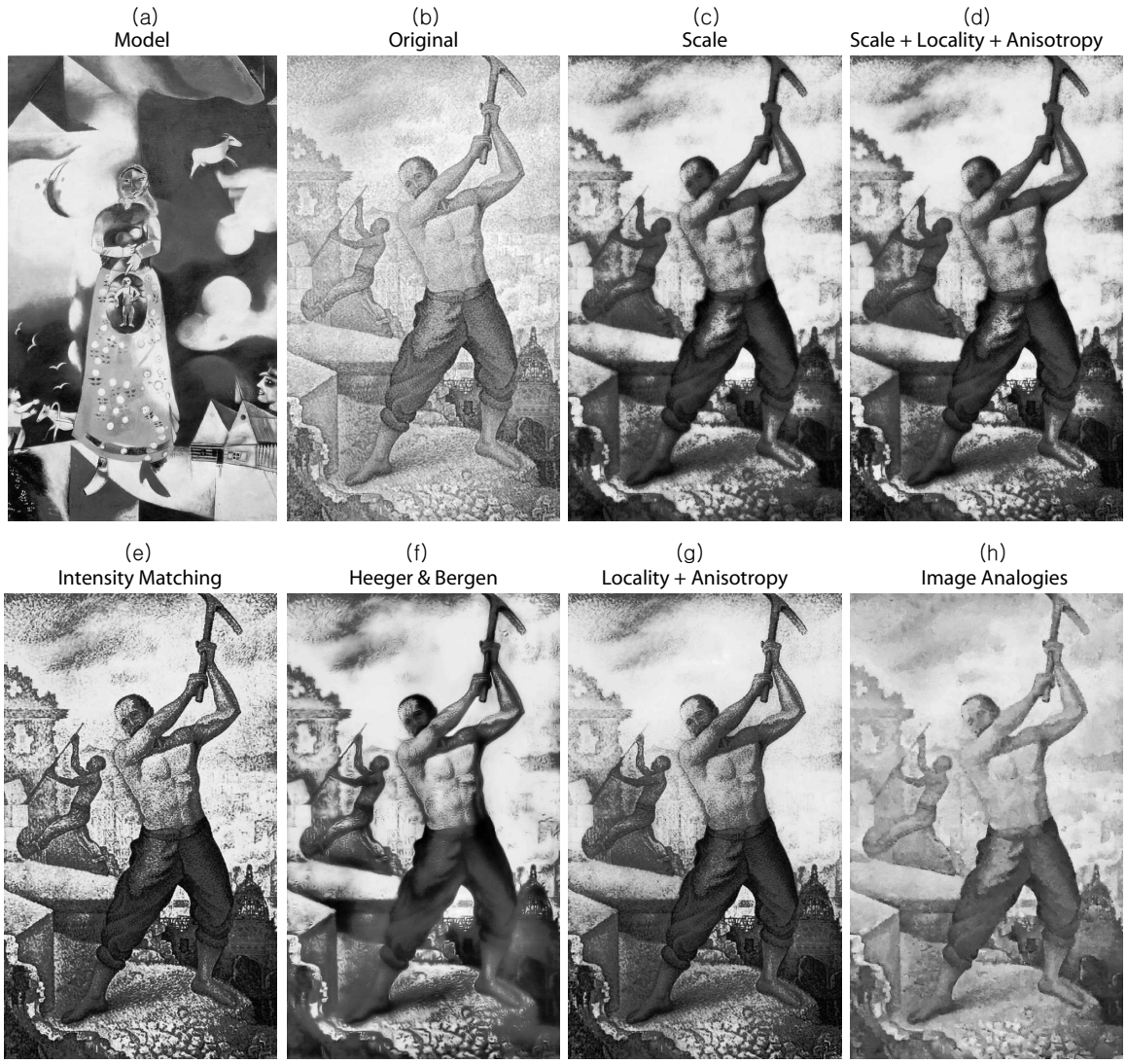


Figure 7-1: Multi-scale transfer from Chagall's *Maternity*(a) to Signac's *Le Demolisseur*(b).

Figure 7-2: Multi-scale transfer from Signac's *Le Demolisseur*(a) to Chagall's *Maternity*(b).

transfer (c), the whole transfer (d), and the non-stationarity and anisotropy transfer (g) do not transfer the point-shaped brush strokes of Pointillism since our method does not capture the correlation of frequency content across spatial domain or across orientations. However, with the transfer of multi-scale features, the scale transfer (c) and the whole transfer (d) decrease the large-scale contrast. This generates a flat look, which is more similar to the model image (a) than the intensity matching (e) and the non-stationarity and anisotropy transfer (g) images are. On the other hand, due to the independent transfer of oriented subbands, the histogram transfer of Heeger and Bergen [25] (f) generates noise effect, which somehow achieves the style of pointillism from the model image (a).

Multi-scale transfer changes the large-scale contrast and the strength of high-frequency content. When the large-scale contrast increases, an image becomes three-dimensional, whereas an image becomes flat with decreased contrast. Decreasing high-frequency content can cause two different effects: texture reduction and line blurring. On the other hand, increasing high-frequency content causes textures and lines to become sharper. However, our transfer does not separate textures and lines. This will be discussed in Section 7.3.

7.2 Non-stationarity Transfer

In the previous section, the multi-scale transfer generated a similar look, but the variation of detail was not transferred. As we discussed in Section 5.2, the distribution of local detail changes the location that receives attention and the depth of field effect that can be used to compensate for the flatness of two dimensional canvas.

To transfer the notion of spatial variation, we take a $k \times k$ size window $W_{s,i}$ in the subband B_s for each pixel x_i and compute the average amplitude across all i, $\overline{|W_{s,i}|}$. We use $\frac{128}{2^{s-1}} + 1$ for k at each scale s, from 1 to 5. We match the histogram of the normalized average amplitude, $\frac{\overline{|W_{s,i}|}}{\overline{|B_s|}}$, of an input image and a model image. HT(a,b)

stands for histogram transfer from a to b.

$$W_{s,i}^{new} = HT(\frac{\overline{|W'_{s,i}|}}{\overline{|B'_s|}}, \frac{\overline{|W_{s,i}|}}{\overline{|B_s|}}), \tag{7.2}$$

where i = spatial location and s = scale

We match two histograms of box-filtered coefficients of steerable pyramid subbands of an image and a model image. We multiply the ratio of $W_{s,i}^{new}$ to $\frac{\overline{|W_{s,i}|}}{\overline{|B_s|}}$ to the pixel i at the scale s. After this non-stationarity transfer, we apply pixel histogram matching, as Heeger and Bergen transfer [25] does.

When we match histograms of the windows, we can transfer the distribution of contrast over an image. In Figure 7-3, the non-stationarity transfer (c) increases the contrast inside the tree and blurs the contrast in the background. This enhances the depth-of-field effect. Note that Hertzmann et al.'s image analogies technique [26] removes Van Gogh's spin-whirl brush strokes, but does not transfer the depth-of-field effect.

In Figure 7-4, the non-stationary transfer (c) increases the depth of the field effect of the same model image to a different image, which is a portrait. The non-stationarity transfer makes the girl in the foreground focused as opposed to the background blurred, which cannot be done by the intensity transfer (e) or by the histogram match of Heeger and Bergen (f).

The non-stationarity transfer makes the regions with lots of details blurred and some regions without high-frequency features sharper. However, the blurring might result in artifacts. In addition, the decision on where to blur or sharpen is based on the initial frequency content. The sharpening might increase noise instead of oriented features such as lines, edges, or brush strokes. Noise in images can be recognized as high-frequency content including JPEG artifacts. Increasing the amplitude of high frequency might generate unpleasant noisy images.

In addition, the non-stationarity aspects of pictorial styles depend on the image contents. Landscape images with a sky tend to have large variation of details in the spatial domain. That is, low average amplitude is present in the sky and higher

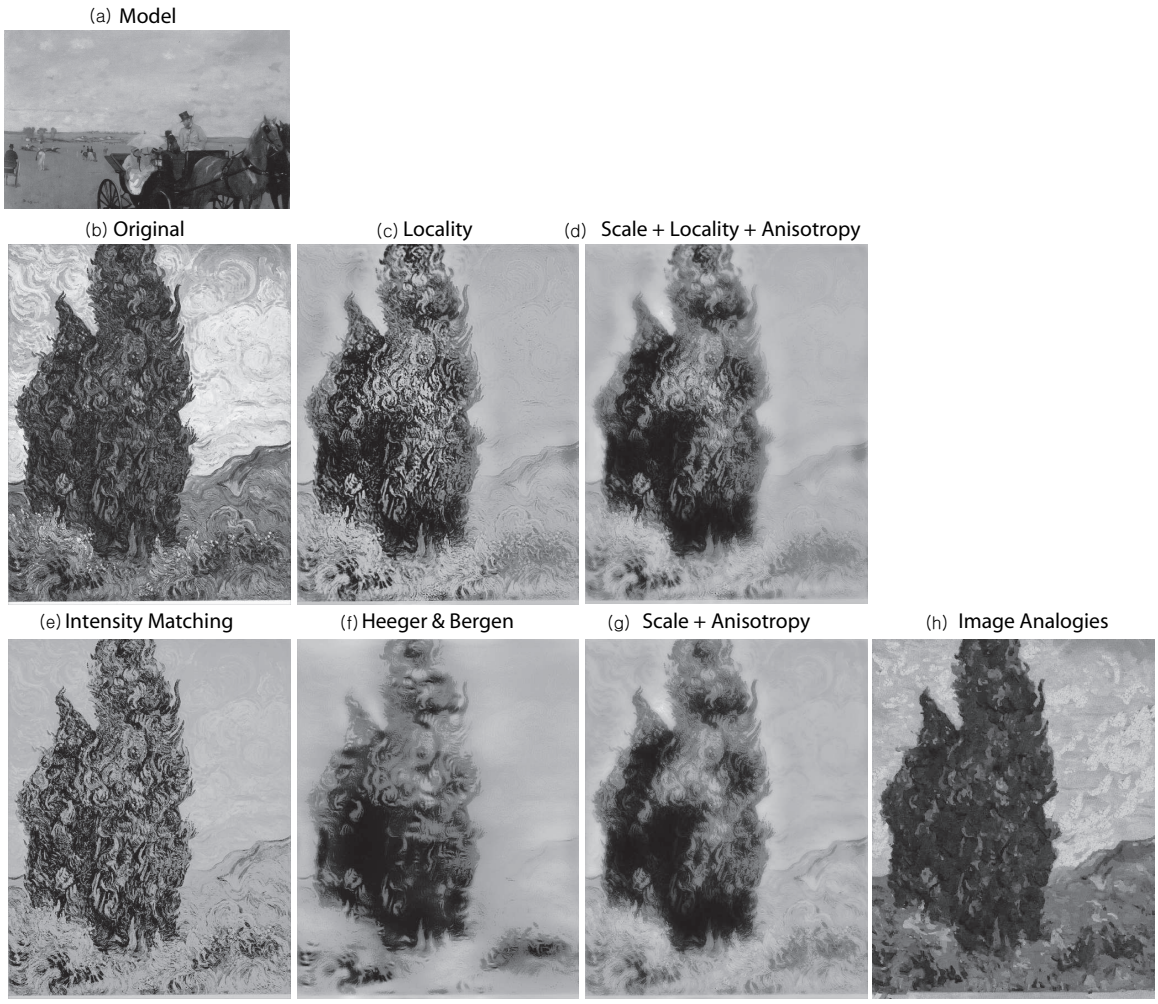


Figure 7-3: Non-stationarity transfer from Degas' *Aux courses en province* to Van Gogh's *Cypresses*. (c) transfers the depth of field effect of (a). The details in the foreground increase, but the high-frequency content in the background becomes blurred.

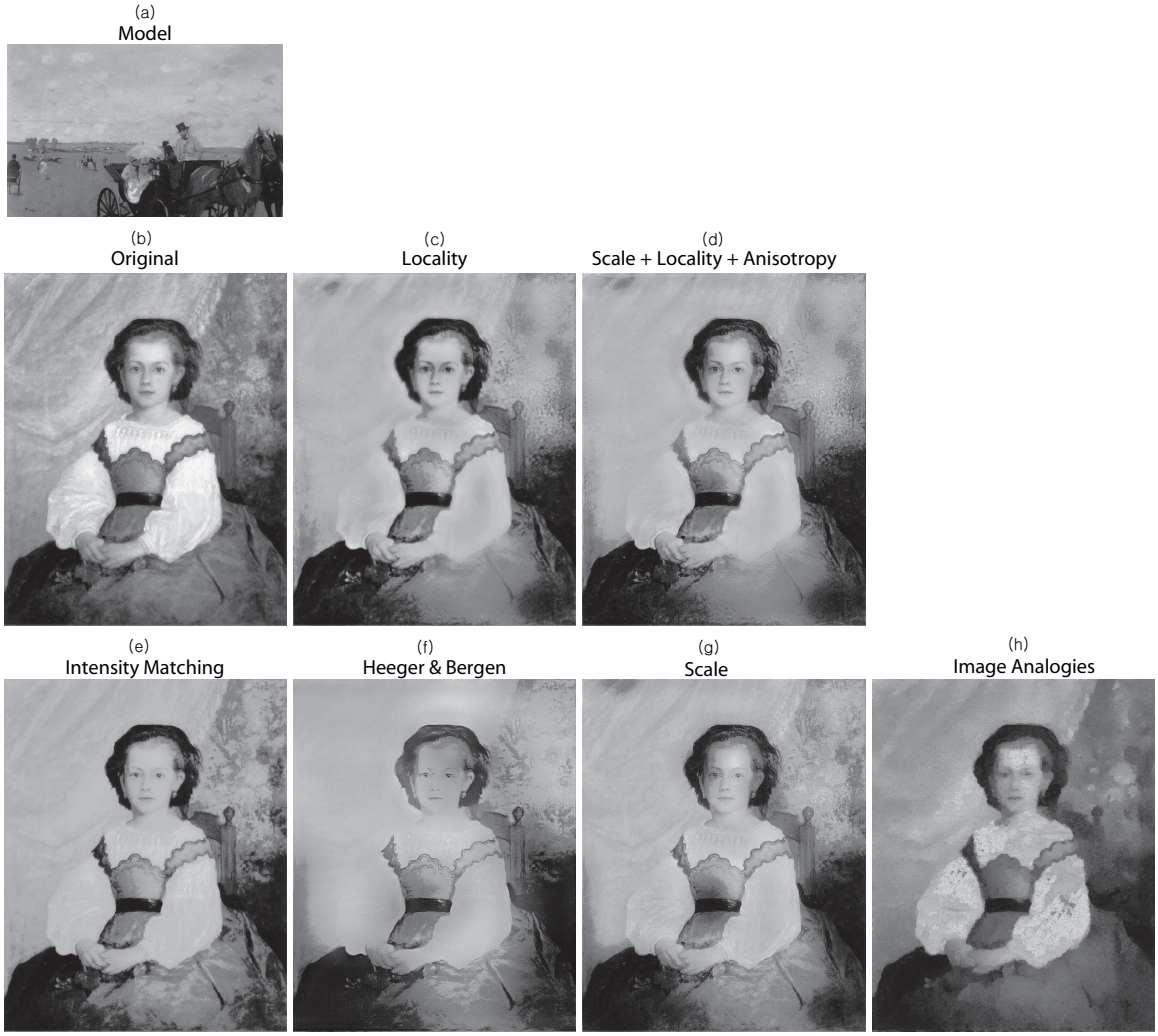


Figure 7-4: Non-stationarity transfer from Degas' *Aux courses en province* to Renoir's *Portrait de Romaine Lacaux*. (c) and (d) transfer the depth of field effect of (a). The high-frequency in the foreground remains, but the high-frequency content in the background becomes blurred. In this portrait, the face becomes focused.

values are generated in the other regions. Portraits have low average amplitude in the background and high average amplitude in the people's faces. Unfortunately, the non-stationarity transfer depends not only on pictorial styles but also on image contents.

7.3 Anisotropy Transfer

In the previous sections, we explored the transfer of the multi-scale and non-stationary features. Mere sharpening and blurring may lead to increased noise or blurred oriented structures such as lines and edges. It is necessary to distinguish frequency content relevant to oriented and non-oriented structure. The standard deviation across orientations distinguishes oriented structures from "noise". Increasing standard deviation performs anisotropic sharpening of edges, while decreasing standard deviation performs anisotropic blurring of edges.

$$\frac{std(|B_{s,k}(i)|)^{new}}{\overline{|B_s|}} = HT(\frac{std(|B'_{s,k}(i)|)}{\overline{|B'_s|}}, \frac{std(|B_{s,k}(i)|)}{\overline{|B_s|}}) \ , \tag{7.3}$$

where s = scale, i = pixel location, and k = orientation

We have the oriented subband $B_{s,k}$ of each subband B_s in the steerable pyramid for each $k\ th$ orientation. At each point x_i, we compute $std(|B_{s,k}(i)|)$, the standard deviation of $B_{s,k}(i)$ across four orientations. Then we match the histogram of the standard deviations over the whole image. We normalize this anisotropy feature by the average amplitude of B_s. According to the new standard deviation, we modify each coefficient. Note that changing the standard deviation changes the maximum and the minimum of coefficient, which effects the total energy of the steerable pyramid subband. After this anisotropy transfer, we apply pixel histogram matching, as Heeger and Bergen transfer [25] does.

Figure 7-5 compares the anisotropy and multi-scale transfer results. The notion of sharp lines was transferred in multi-scale and anisotropy transfer. This might

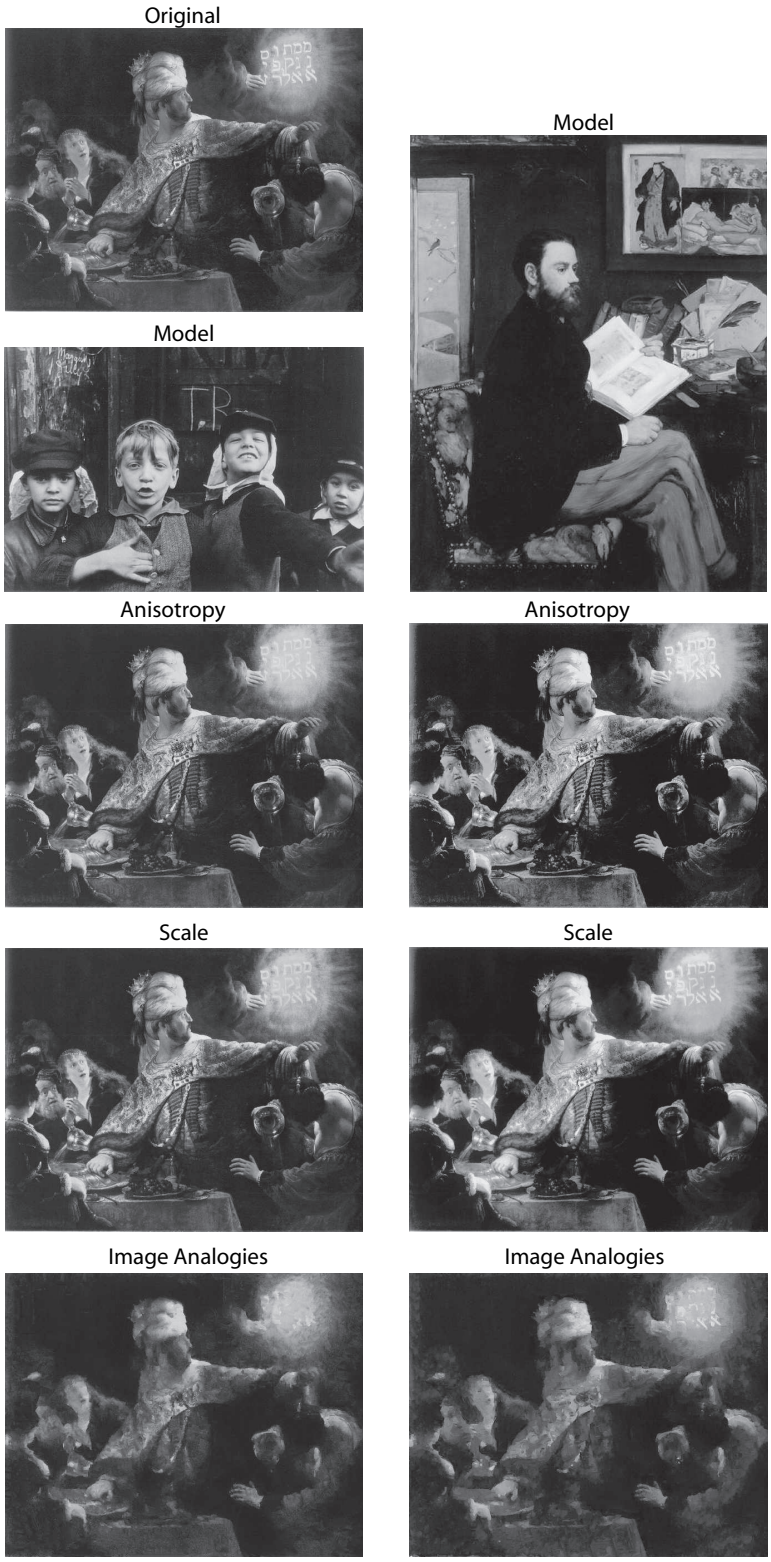


Figure 7-5: Transfer results from Levitt's *New York* to Rembrandt's *The Feast of Belshazzar* are in the first column, and those from Manet's *Portrait d'Emile Zola* to Rembrandt's *The Feast of Belshazzar* are in the second column. Anisotropy and multi-scale transfer results generate similar effects, but the anisotropy transfer generates less noise.

suggest that they are correlated. However, the anisotropy transfer generates more refined edges and creates less noise. This anisotropy aspect regarding edges is not transferred by Hertzmann et al.'s image analogies [26] for both model images in figure 7-5. In general, Hertzmann et al.'s image analogies [26] are not suitable to transfer photographic styles. Figure 7-6 shows another example.

Anisotropy features capture oriented structures separate from non-oriented structures of pictorial styles. When we combine the multi-scale and anisotropy transfer, the anisotropy transfer effect is not visible. Figure 7-6 shows such an example. The multi-scale transfer dominates the visual results. In the previous chapter, the result of LDA showed that the anisotropy aspects of pictorial styles are important to separate paintings by two different artists. This suggests that we should increase the weight for the effect of the transfer of the anisotropy. We will explore this scaling issue in future work.

7.4 Comparison

In this section, we compare our results with other parametric transfer methods such as Portilla and Simoncelli's texture analysis and synthesis [50] and Heeger and Bergen's histogram matching [25] even though they were not developed for style transfer. We also provide comparison with Hertzmann et al.'s image analogies [26].

Figure 7-7 shows that Portilla and Simoncelli's texture analysis and synthesis [50] is not adequate for the purpose of style analysis and transfer. It does not separate the image content from pictorial style and tends to mimic the structure of objects in the images. It imposes image structures using joint statistics and it does not preserve image content, when transferring between two images. (See Figure 7-7.)

In contrast, Heeger and Bergen's histogram transfer [25] does not change image content such as lines, edges, or composition. The histogram matching modifies the global frequency, whereas our method modifies the distribution of details and the sharpness of oriented structures. Even though Heeger and Bergen's histogram transfer [25] generates convincing results, it does not consider non-stationarity of image

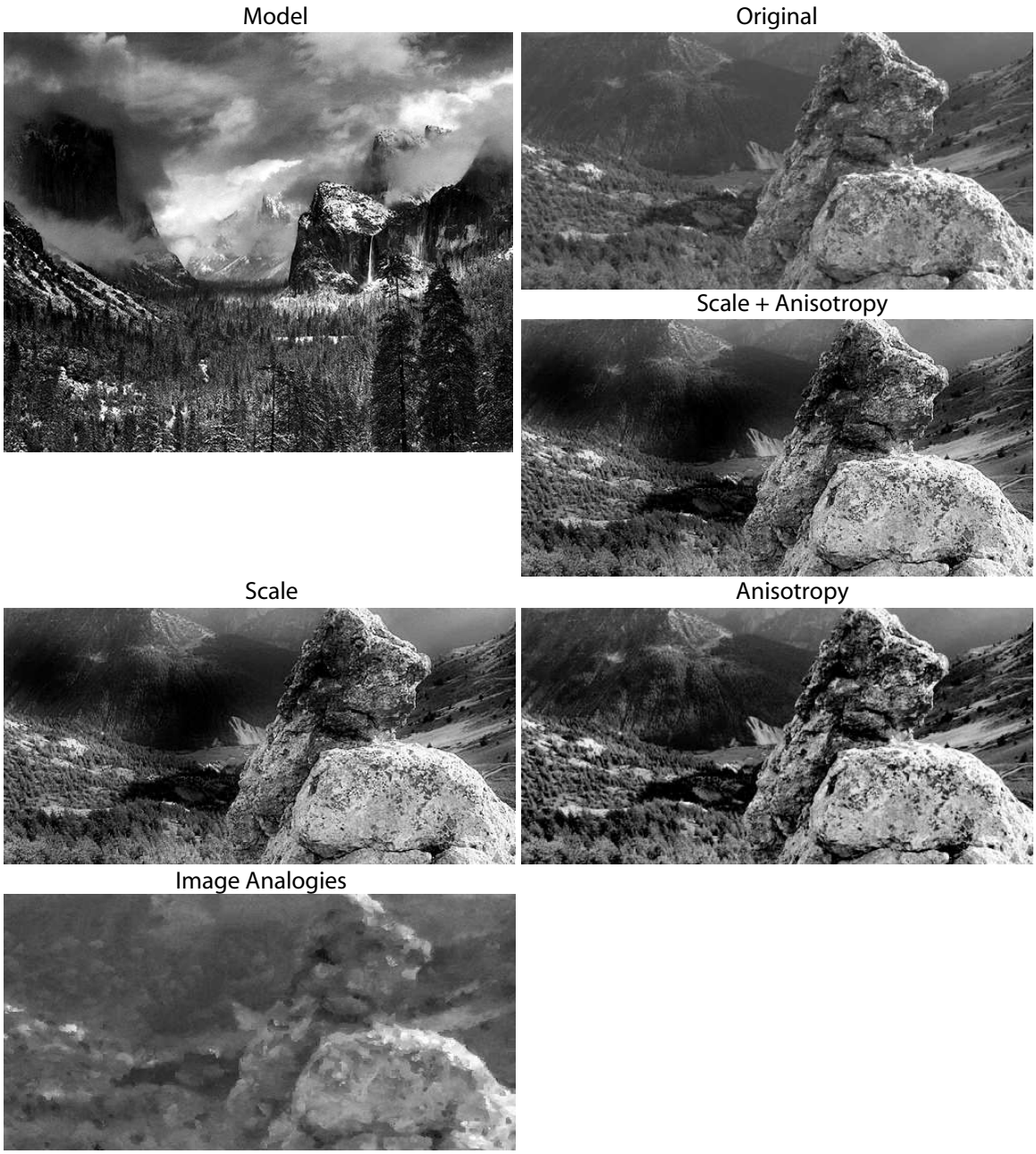


Figure 7-6: Transfer from Ansel Adams' *Clearing Winter Storm* to a natural image. Anisotropy and multi-scale transfer results generate similar effects, but the anisotropy transfer generates less noise. When we combine the multi-scale and anisotropy transfer, the anisotropy transfer effect is not visible.

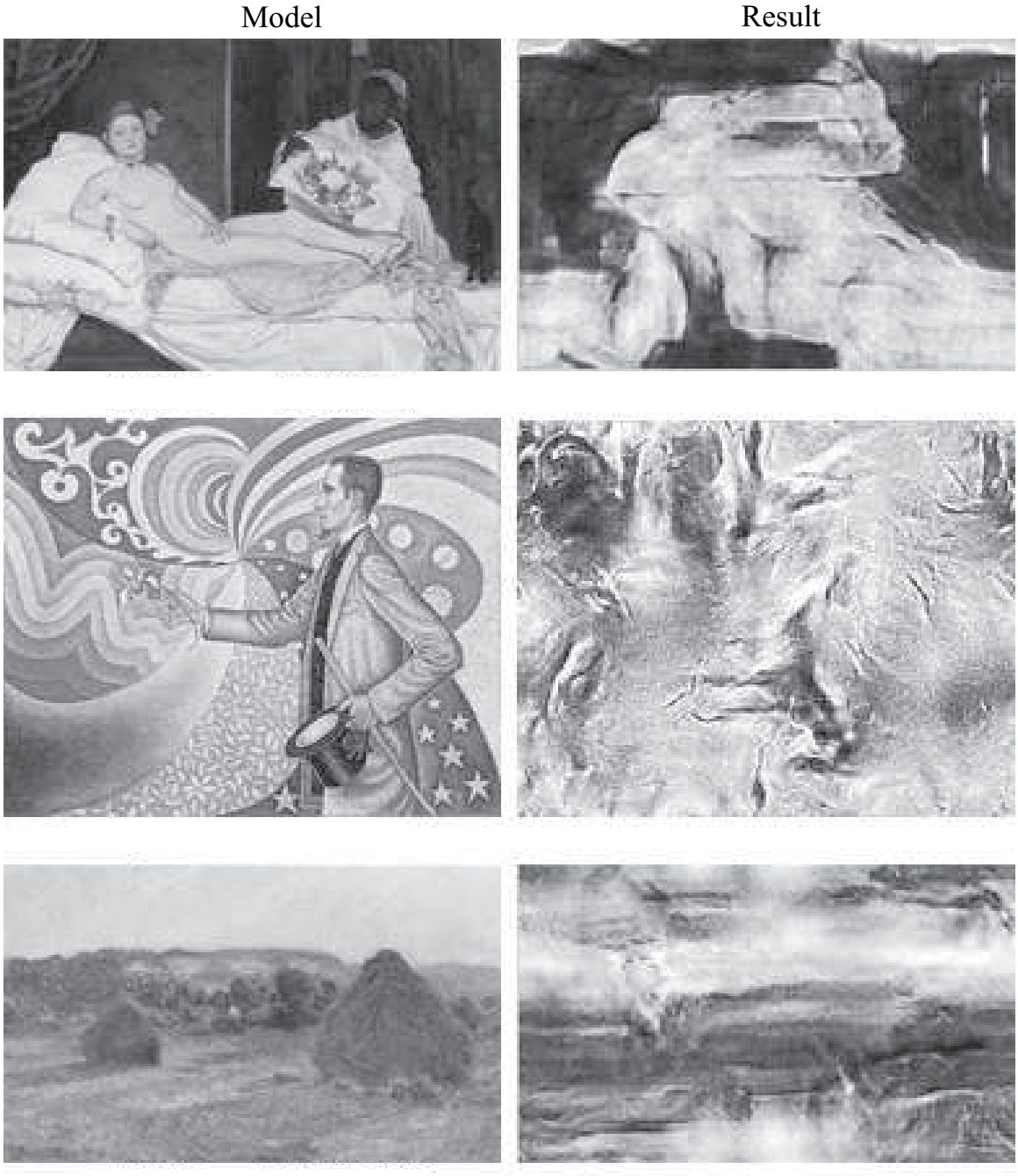


Figure 7-7: The result of Portilla and Simoncelli's texture analysis and synthesis. [50]. Model images are Manet's *Olympia*, Signac's *Against the Enamel of a Background Rhythmic with Beats and Angles, Tones and Colors, Portrait of M. Felix Feneon in 1890*, and Monet's *Wheatstacks (End of Summer)*.

statistics and anisotropy of pictorial styles.

Hertzmann et al.'s image analogies [26] capture fine-grain shapes of brush strokes, but do not transfer large-scale contrast, depth-of-field effect, and anisotropy aspects regarding edges.

Luminance transfer in Figure 7-1, 7-2, 7-3, 7-4, and 7-8 generates convincing results, but it does not change frequency aspects of pictorial styles.

Figure 7-8 shows the difference between the four transfer methods. The intensity transfer does not capture the frequency aspects of pictorial styles. Heeger and Bergen's histogram transfer captures the global frequency aspects of pictorial styles, but does not change the distribution of details or the anisotropy of lines and edges. The results of Hertzmann et al.'s image analogies [26] do not capture differences of pictorial styles. The last row of Figure 7-8 was added as a sanity check and shows that the transfer methods leave the image the same when the model image is a similar image from the same artist as the input image. Compared to these, our method transfers large-scale contrast, depth-of-field effect, and anisotropy aspects regarding edges.

Figure 7-8: Comparison with the intensity matching, Heeger and Bergen's histogram transfer [25], Hertzmann et al.'s image analogies. [26], and our method.

Chapter 8

Discussion

In this thesis, we proposed a new methodology for the quantitative characterization of coarse grain pictorial style. The heart of our approach is the use of both analysis and transfer with the same set of image features. We have shown that the statistical features of paintings and photographs in the frequency domain are related to simple notions of pictorial style. We have validated the relevance of the features to pictorial styles using machine learning techniques. In addition, we transfer pictorial styles between two images by matching the statistical features, which is a crucial complementary validation tool for our features.

Our main contribution is that we analyze and transfer pictorial styles at the same time. Most research on image analysis has not developed tools to transfer statistical regularities and differences. In addition, although we do not provide fine-grain imitation as recent non-parametric transfer methods do, our method allows statistical analysis and a direct manipulation and interpolation of pictorial style.

We have characterized and transferred three aspects of pictorial styles: multi-scale, non-stationarity, and anisotropy aspects. Our features do not generate a clear boundary between artists and do not reproduce texture details like image analogies [26]. Instead, our method captures coarse-grain aspects of pictorial styles. We characterize the large-scale contrast and the distribution of details, which are relevant to pictorial styles, especially pictorial composition.

Our parametric model has more applications beyond transfer. Our features can be

used for direct interpolation, manipulation and enhancement of digital images such as tone mapping. In addition, we want to extend our analysis and transfer to drawing, which includes effects such as pure line drawing and hatching.

In our future work, we want to refine our features. Our current feature set does not separate edges from structures such as textures. This can be improved by studying joint statistics, which can examine the coherence of steerable pyramid coefficients at consecutive scales and in the adjacent spatial location. We seek to separate oriented structures for analysis and transfer.

In addition, manipulation of the steerable pyramid does not preserve edges and lines, and suffers from haloing artifacts due to its band-limited linear basis. We hope to find different image transforms that characterize and transfer pictorial styles without artifacts and that is suitable to separate oriented structures such as edges and lines.

As we refine our feature set, we hope our statistical model can be used for more precise style classification and retrieval. For this end, we need to explore scaling issues. The scale between different types of features should be adjusted properly. In addition, we want to explore different machine learning techniques such as nearest neighbor or neighborhood components analysis to understand correlations and dependencies of our features.

Bibliography

[1] Mark Harden's Artchive. http://www.artchive.com/.

[2] H. Barlow. The exploitation of regularities in the environment by the brain. *Behavioral and Brain Science (BBS)*, 24(4):602–607, 2001.

[3] Kobus Barnard and David A. Forsyth. Learning the semantics of words and pictures. In *ICCV*, pages 408–415, 2001.

[4] Robert W. Buccigrossi and Eero P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. In *IEEE Trans. Image Proc.*, volume 8, pages 1688–1701, 1997.

[5] Chad Carson, Megan Thomas, Serge Belongie, Joseph M. Hellerstein, and Jitendra Malik. Blobworld: A system for region-based image indexing and retrieval. In *Third International Conference on Visual Information Systems*. Springer, 1999.

[6] F. Cutzu and R. Hammoud. Estimating the photorealism of images: Distinguishing paintings from photographs. In *Proc. of CVPR*, 2003.

[7] J. Daugman. Two-dimensional spectral analysis of cortical receptive field profile. In *Vision Research*, volume 20, pages 847–856, 1980.

[8] J. Daugman. Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two-dimensional visual cortical filters. In *J. Opt. Soc. Am.*, volume 2, pages 1160–1169, 1985.

[9] J. Daugman. Complete discrete 2d gabor transforms by neural networks for image analysis and compression. In *IEEE Transactions on Acoustics, Speech and Signal Processing*, volume 36, pages 1169–1179, 1997.

[10] Ron O. Dror, Edward H. Adelson, and Alan S. Willsky. Recognition of surface reflectance properties from a single image under unknown real-world illumination. In *Proc. IEEE workshop on IOAV in Lighting : Psychophysics and Computation*, December 2001.

[11] Ron O. Dror, Thomas K. Leung, Edward H. Adelson, and Alan S. Willsky. Statistics of real-world illumination. In *Proceedings of CVPR*, December 2001.

[12] Iddo Drori, Daniel Cohen-Or, and Hezy Yeshurun. Example-based style synthesis. In *Proc. of CVPR*, 2003.

[13] Shlomo Dubnov, Naftali Tishby, and Dalia Cohen. Hearing beyond the spectrum. *Journal of New Music Research*, 24(4), 1995.

[14] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *International Conference on Computer Vision*, page 1033?1038, 1999.

[15] Alexei A. Efros and William T. Freeman. Image quilting for texture synthesis and transfer. In *Proc. SIGGRAPH*, 2001.

[16] Fairchild. *Color Appearance Models*. Addison-Wesley, 1998.

[17] Hany Farid and Siwei Lyu. Higher-order wavelet statistics and their application to digital forensics. In *IEEE Workshop on Statistical Analysis in Computer Vision (in conjunction with CVPR)*, 2003.

[18] G.T. Fechner. *Elemente der Psychophysik I u. II.* Breitkopf u. Hartel, Leipzig, 1907.

[19] David J. Field. Relations between the statistics of natural images and the response properties of cortical cells. 4(12):2379–2394, December 1987.

[20] D.J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12):2379–2394, 1987.

[21] Myron Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele, and Peter Yanker. Query by image and video content: The qbic system. *Computer*, 28(9):23–32, 1995.

[22] W. Freeman, J. Tenenbaum, and E. Pasztor. An example-based approach to style translation for line drawings. Technical Report 99-11, MERL, 1999.

[23] William T. Freeman and Edward H. Adelson. The design and use of steerable filters. *IEEE Trans. Patt. Anal. and Machine Intell.*, 13(9), 1991.

[24] E. Gombrich. *The Story of Art*. Phaidon Press Ltd, 1950.

[25] David J. Heeger and James R. Bergen. Pyramid-based texture analysis/synthesis. In *Proc. SIGGRAPH*, 1995.

[26] Aaron Hertzmann, Charles E. Jacobs, Nuria Oliver, Brian Curless, and David H. Salesin. Image analogies. In *Proc. SIGGRAPH*, 2001.

[27] Aaron Hertzmann, Nuria Oliver, Brian Curless, and Steven M. Seitz. Curve analogies. In *EG Workshop on Rendering*, 2002.

[28] Jinggang Huang, Ann B. Lee, and David Mumford. Statistics of range images. In *Proc. CVPR*, 2000.

[29] Jinggang Huang and David Mumford. Statistics of natural images and models. In *Proc. Conf. CVPR*, pages 541–7, 1999.

[30] O. Icoglu, B. Gunsel, and S. Sariel. Classification and indexing of paintings based on art movements. In *12th European Signal Processing Conference (Eusipco)*, pages 749–752, September 2004.

[31] T. Jehan and B. Schoner. An audio-driven perceptually meaningful timbre synthesizer. In *Proceedings International Computer Music Conference*, 2001.

[32] Robert D. Kalnins, Lee Markosian, Barbara J. Meier, Michael A. Kowalski, Joseph C. Lee, Philip L. Davidson, Matthew Webb, John F. Hughes, and Adam Finkelstein. Wysiwyg npr: Drawing strokes directly on 3d models. *ACM Trans. on Graphics*, 21(3), 2002.

[33] D. Keren. Recognizing image "style" and activities in video using local features and naive bayes. *Pattern Recognition Letters*, 2003. forthcoming.

[34] Daniel Keren. Painter identification using local features and naive bayes. In *ICPR (2)*, pages 474–477, 2002.

[35] Daniel Keren. Recognizing image "style" and activities in video using local features and naive bayes. *Pattern Recognition Letters*, 24(16):2913–2922, 2003.

[36] John-Peter Lewis. Texture synthesis for digital painting. In *Computer Graphics (Proceedings of SIGGRAPH 84)*, volume 18, pages 245–252, July 1984.

[37] A. Leykin and F. Cutzu. Distinguishing paintings from photographs. *Journal of Vision*, 2(7), 2002.

[38] A. Leykin and F. Cutzu. Differences of edge properties in photographs and paintings. In *IEEE International Conference on Image Processing*, 2003.

[39] Jia Li, Robert M. Gray, and Richard A. Olshen. Multiresolution image classification by hierarchical modeling with two-dimensional hidden markov models. *IEEE Transactions on Information Theory*, 46(5):1826–1841, 2000.

[40] Jia Li and James Z. Wang. Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(9):1075–1088, 2003.

[41] Jia Li and James Z. Wang. Studying digital imagery of ancient paintings by mixtures of stochastic models. *IEEE Trans. Image Processing*, 12, 2004. accepted for publication.

[42] Jia Li, James Ze Wang, and Gio Wiederhold. IRM: integrated region matching for image retrieval. In *ACM Multimedia*, pages 147–156, 2000.

[43] B. Logan and A. Salomon. A music similarity function based on signal analysis. In *ICME 2001.*, 2001.

[44] Thomas Lombardi, Sung-Hyuk Cha, and Charles Tappert. A graphical user interface for a fine-art painting image retrieval system. In *MIR '04: Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*, pages 107–112. ACM Press, 2004.

[45] T. Melzer, P. Kammerer, and E. Zolda. Stroke detection of brush strokes in portrait miniatures using asemi-parametric and a model based approach. In *ICPR '98: Proceedings of the 14th International Conference on Pattern Recognition-Volume 1*, pages 474–476. IEEE Computer Society, 1998.

[46] Sumanta N. Pattanaik, James A. Ferwerda, Mark D. Fairchild, and Donald P. Greenberg. A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of SIGGRAPH 98*, Computer Graphics Proceedings, Annual Conference Series, pages 287–298, Orlando, Florida, July 1998. ACM SIGGRAPH / Addison Wesley.

[47] A. Pentland, R. W. Picard, and S. Sclaroff. Photobook: content-based manipulation of image databases. *Int. J. Comput. Vision*, 18(3):233–254, 1996.

[48] S. Polyak. *The Vertebrate Visual System.* Univ. Chicago Press, Chicago, 1957.

[49] Javier Portilla and Eero P. Simoncelli. Image denoising via adjustment of wavelet coefficient magnitude correlation. In *Proc. of the 7th Int'l Conf. on Image Proc.*, September 200.

[50] Javier Portilla and Eero P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *IJCV*, 40(1), October 2000.

[51] Javier Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. In *IEEE Transactions on Image Processing*, volume 12, pages 1338– 1351, November 2003.

[52] Erik Reinhard, Peter Shirley, Michael Ashikhmin, and Tom Troscianko. Second order image statistics in computer graphics. In *APGV '04: Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pages 99–106, New York, NY, USA, 2004. ACM Press.

[53] Erik Reinhard, Peter Shirley, and Tom Troscianko. Natural image statistics for computer graphics. Technical Report UUCS-01-002, University of Utah, March 2001.

[54] Daniel L. Ruderman. The statistics of natural images. *Network: Comput. Neural Syst.*, 5:517–548, 1994.

[55] Daniel L. Ruderman and W. Bialek. The statistics of natural images : Scaling in the woods. *Physics Review Letter*, 73(6):814–817, 1994.

[56] Y. Rui, T. Huang, and S. Chang. Image retrieval: current techniques, promising directions and open issues, April 1999.

[57] Robert Sablatnig, Paul Kammerer, and Ernestine Zolda. Hierarchical classification of paintings using face- and brush stroke models. In *ICPR '98: Proceedings of the 14th International Conference on Pattern Recognition-Volume 1*, page 172. IEEE Computer Society, 1998.

[58] E P Simoncelli and B A Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24:1193–1216, May 2001.

[59] Eero P. Simoncelli. Statistical models for images: Compression, restoration and synthesis. In *31st Asilomar Conf. on Signals, Systems and Computes*, pages 673–678, November 1997.

[60] Eero P. Simoncelli. Modeling the joint statistics of images in the wavelet domain. In *Proc. SPIE*, volume 3813, 1999.

[61] Eero P. Simoncelli and Willian T. Freeman. The steerable pyramid : A flexible architecture for multi-scale derivative computation. In *Second Int'l Conf. on Image Proc.*, volume 3, pages 444–447, October 1995.

[62] Eero P. Simoncelli, Willian T. Freeman, Edward H. Adelson, and David J. Heeger. Shiftable multi-scale transforms. 38(2):587–607, March 1992.

[63] Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380, 2000.

[64] Antonio Torralba and Aude Oliva. Statistics of natural image categories. *Network: Comput. Neural Syst.*, 14:391–412, January 2003.

[65] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5), 2002.

[66] Brian A. Wandell. *Foundations of Vision*. Sinauer Associates, Inc., 1995.

[67] James Z. Wang, Jia Li, and Ching chih Chen. Interdisciplinary research to advance digital imagery indexing and retrieval technologies for asian art and cultural heritages. In *Proc. 4th International Workshop on Multimedia Information Retrieval, in conjunction with ACM Multimedia*, 2002.

[68] James Ze Wang, Jia Li, and Gio Wiederhold. SIMPLIcity: Semantics-sensitive integrated matching for picture LIbraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):947–963, 2001.

[69] E.H. Weber. *De pulsu, resorptione, audita et tactu. -Annotationes anatomicae et physiologicae-*. Trs. by H.E. Ross, Academic Press, New York, 1978, 1834.

[70] M. Welsh, N. Borisov, J. Hill, R. von Behren, and A. Woo. Querying large collections of music for similarity. Technical Report UCB/CSD00 -1096, UC Berkeley CS Div., 1999.

[71] D. Wessel. Timbre space as a musical control structure. *Computer Music Journal*, 3(2), 1979.

[72] Gerhard Widmer, Simon Dixon, Werner Goebl, Efstathios Stamatatos, and Asmir Tobudic. Empirical music performance research: OFAI's position. Panel Discussion Paper, MOSART Workshop on Current Research Directions in Computer Music, 2001.

[73] M. Wright, A. Chaudhary, A. Freed, S. Khoury, and D. Wessel. Audio applications of the sound description interchange format standard, 1999.

[74] James Z. Wang Yixin Chen and Robert Krovetz. Clue: Cluster-based retrieval of images by unsupervised learning. In *IEEE Transactions on Image Processing*, volume 13, 2004.