AUTOMATED ANALYSIS OF COMPOSITION AND STYLE OF PHOTOGRAPHS AND PAINTINGS

A Dissertation in
Information Sciences and Technology

by
Lei Yao

© 2013 Lei Yao

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2013
The dissertation of Lei Yao was reviewed and approved* by the following:

James Z. Wang  
Professor of Information Sciences and Technology  
Dissertation Co-advisor, Co-chair of Committee

Jia Li  
Professor of Statistics  
Dissertation Co-advisor, Co-chair of Committee

David Hall  
Dean in the College of Information Sciences and Technology  
Professor of Information Sciences and Technology

Xiaolong (Luke) Zhang  
Associate Professor of Information Sciences and Technology

Patrick J. McGrady  
Charles V. Hallman Curator, Palmer Museum of Art  
Affiliate Assistant Professor of Art History

Madhu C. Reddy  
Graduate Program Director  
Associate Professor of Information Sciences and Technology

*Signatures are on file in the Graduate School.
Abstract

Computational aesthetics is a newly emerging cross-disciplinary field with its core situated in traditional research areas such as image processing and computer vision. Using a computer to interpret aesthetic terms for images is very challenging. In this dissertation, I focus on solving specific problems about analyzing the composition and style of photographs and paintings. First, I studied the problem of distinguishing van Gogh’s paintings from his contemporaries. The application of rhythmic and spontaneous brushstrokes is a prominent trait for van Gogh’s paintings. His unique brushstroke style is characterized by features calculated from automatically extracted brushstrokes. Statistical analysis on the extracted brushstroke features shows success in tackling real-world painting analysis tasks designed by art historians. Second, I explore the possibility of characterizing styles of paintings without visible brushstrokes, specifically artworks by Norval Morrisseau. Curve elegance measurements are used to differentiate authentic works from forgeries. Then, I present my studies on the topic of photography composition. Composition is closely related to the aesthetic qualities of images and a key factor that distinguishes professional photographs from snapshots. I design a spatial composition classifier to analyze the compositional properties for general photographs. A new integrated system is presented to render on-site photography feedback for users by retrieving high-quality exemplar photographs with similar compositions. User studies substantiate the system’s performance. Finally, I propose a dark-light re-composition algorithm to emulate the dodging and burning techniques used in darkroom photography. The algorithm performs region-wise intensity adjustments by utilizing the intensity distribution and the Notan structure of an exemplar. Overall, this dissertation studies computational approaches to the interpretation of artistic terms and explores potential applications in digital painting analysis, automatic photography feedback and photo enhancement.
# Table of Contents

List of Figures vii  
List of Tables x  
Acknowledgments xi  

Chapter 1  
Introduction 1  
1.1 Contributions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5  
1.2 Organization of this Dissertation . . . . . . . . . . . . . . . . . .. 6  

Chapter 2  
Computational Aesthetics: Painting Analysis, Photograph Assessment and Enhancement 7  
2.1 Painting Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8  
2.2 Aesthetics Assessment and Enhancement for Photographs . . . . . . 11  
2.2.1 Aesthetics Assessment . . . . . . . . . . . . . . . . . . . . . . 12  
2.2.2 Aesthetics Enhancement by Cropping and Retargeting . . . . . 13  
2.2.3 Exemplar-based Style Transfer . . . . . . . . . . . . . . . . . 14  

Chapter 3  
Distinguish van Gogh from his contemporaries via Automated Brushstroke Extraction 17  
3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17  
3.2 Brushstroke Extraction . . . . . . . . . . . . . . . . . . . . . . . . 21  
3.2.1 Algorithm Overview . . . . . . . . . . . . . . . . . . . . . . . 22  
3.2.1.1 Edge Detection Based Brushstroke Extraction . . . . . 22  
3.2.1.2 Image Segmentation Based Brushstroke Extraction 24
## List of Figures

1.1 Paradigm for computer vision research. ........................................ 2

3.1 van Gogh’s paintings versus his contemporaries’. .......................... 18
3.2 van Gogh’s painting from Paris period. ....................................... 19
3.3 van Gogh’s painting from Arles-Saint-Rémy period. ......................... 20
3.4 Three paintings that receive different opinions on dating. ................. 20
3.5 Flowchart of the brushstroke extraction algorithm. .......................... 22
3.6 Brushstroke extraction based on edge detection. (a) A painting patch.  (b) Edge pixels detected by the edge detection algorithm in [55]. (c) After short edges (less than 30 pixels) are removed. (d) After enclosing small edge gaps (distance less than 15). ............... 23
3.7 Brushstroke extraction based on image segmentation. Left: image segmentation results. Right: pixels that belong to brushstrokes extracted in the first pass removed. ........................................ 24
3.8 Backbone extraction. (a) A branched skeleton. (b) Two edge segments after edge linking. (c) Two artificial endpoints (in red). (d) Edge segments after inserting artificial endpoints. (e) Backbone. .................. 26
3.9 Examples for extracting backbones in branched skeletons. First row: an example for “not severely branched” skeletons. Second row: an example for a “severely branched” skeleton. (a) Connected components. (b) The skeletons shown as edge segments formed by edge linking. (c) Edge segments formed after inserting artificial endpoints. (d) Edge segments left after merging at branching positions. ........................................ 27
3.10 Brushstroke extraction results for van Gogh paintings. Painting images courtesy of the Van Gogh Museum Amsterdam (Vincent van Gogh Foundation). ........................................ 28
3.11 Comparison of automatically extracted brushstrokes and manually marked brushstrokes. (a) Painting patch. (b) Automatically extracted brushstrokes. (c) Manually marked brushstrokes. Painting IDs from top to bottom: F218, F248a, F297, F374, F386. ............... 31
3.12 After the extraction of brushstrokes, numerical features can be computed. The skeletons and boundaries of the brushstrokes are shown.

4.1 Example authentic and counterfeit paintings from the dataset related to Norval Morrisseau, a native aboriginal painter from Canada.

4.2 Curve detection and linking applied to a painting.

4.3 The edge map of a counterfeit painting.

4.4 Edge points under different conditions.

4.5 Calculated measures for the 19 Morrisseau paintings and 16 forgeries. Values in (a)-(c) are sorted.

5.1 Photographs falls into the five composition categories. Top row from left to right: "vertical", "horizontal", "centered". Bottom row: "diagonal" (diagonal direction from upper left to bottom right), "diagonal" (diagonal direction from upper right to bottom left), "textured".

5.2 Photographs of diagonal composition.

5.3 Diagonal orientation bounding conditions.

5.4 Division of the image into 8 angular areas with respect to a reference pixel.

5.5 Photographs classified into multiple categories. "diagonal_ulbr" represents the diagonal from the upper left corner to the bottom right corner, and "diagonal_urbl" represents the other.

5.6 Images retrieved by different values of $K$, using composition categorization to rerank results returned by SIMPLIcity. (a) The query image is at the top center and top 8 re-ranked images retrieved when $K = 50$ (first row), $K = 100$ (second row) and $K = 300$ (third row) are shown in three rows; (b) A second example with the query image at the top center and the top 8 re-ranked images retrieved when $K = 50$ (first row), $K = 100$ (second row) and $K = 200$ (third row) are shown in three rows.

5.7 Merging edges or segments to form more prominent diagonal elements. First column: original images; Second column: edge detection or segmentation results; Third column: edges or segments that are merged and identified as diagonal elements.

5.9 Example photos classified as “horizontal”, “vertical”, and “centered”. First row: original images. Second row: segmentation result. 68
5.10 Photo examples mistakenly classified as “Centered” by our algorithm. (a) Photos labeled “Vertical” by users; (b) Photos labeled “Diagonal” by users. 72
5.11 Histogram of the number of mismatches between classification by our algorithm and by the users. 73
5.12 Prototype of the OSCAR system. 74
5.13 Homepage of OSCAR demo website. 75
5.14 A potential visual interface for the proposed photography feedback system in mobile applications. 75

6.1 Tone mapping curves with various parameters. 81
6.2 Left to right: the source image, the exemplar image, the intensity histograms (gray for the source image and blue for the exemplar). 83
6.3 Comparison between global and region-wise adjustments. 84
6.4 (a) The source image; (b) The exemplar; (c) Modified image generated by (6.5); (d) Modified image by (6.8); Modified images by (e) global histogram matching and (f) color normalization. 86
6.5 Modification by different Notan structures. 87
6.6 Fuzzy region map. From left to right: original image, image segment and its fuzzy map. 89
6.7 (a) The source image, and (b) a modified version in Photoshop. 90
6.8 Modified images and their histograms. From left to right: exemplar photographs, output by global histogram matching, color normalization and the proposed algorithm. 91
6.9 More experiment results. (a) the source image; (b) the exemplar; (c) results by global histogram matching; (d) results by color normalization [71]; (e) results by the proposed approach. 92
List of Tables

3.1 Comparing Automatically Extracted and Manually Marked Brush-strokes in 10 Patches Cropped from Paintings in the van Gogh Collection. ................................................................. 30
3.2 P-values for the eleven attributes under the study that compares paintings by van Gogh and his contemporaries. ................................. 37
3.3 P-values for the eleven attributes under the study that compares van Gogh’s paintings in Paris period and in Arles-Saint-Rémy period. 38

4.1 Results of 2-sample t-test .................................................. 46
4.2 SVM cross-validation error rates ......................................... 46

5.1 Confusion table for composition classification on 222 images from five categories ................................................................. 69
5.2 Distribution of the entropy for the votes of users. For each composition category, the percentage of photos yielding a value of entropy in any bin is shown. ......................................................... 70
5.3 Confusion table for exclusive classification of 494 images into six composition categories. Each row corresponds to a ground truth class. ................................................................. 71
Acknowledgments

Completing my Ph.D. is thus far the most challenging task in my life. It has been a long journey, and much has happened and changed during those years. Without the invaluable support of many people, I would not be where I am today.

My deep gratitude goes first to my advisors, Professors James Wang and Jia Li, who helped me get started on my degree and guided me through the last six years of graduate school. Thanks to them, I had the chance to explore novel knowledge through cutting-edge research problems. Owing to their guidance, supervision and help, I have learned an enormous amount of knowledge through challenging yet valuable experiences. I am also grateful to them for giving me precious advice in order to improve myself and become a better person.

I would like to thank the members of my dissertation committee, Professors David Hall, Luke Zhang, and Patrick McGrady, for their expertise, accessibility, and suggestions in general. Their recommendations have all been helpful in clarifying and shaping my research goal.

I also owe special thanks to Ella Hendriks for her help with the project on computational analysis of van Gogh’s paintings. Her expertise in Art History has added treasured value to the work.

I would like to thank all my fellow students in the research group for valuable discussions and friendly support throughout these years. In particular, I would like to thank my colleagues Mu Qiao, Poonam Suryanarayan, David Zhang, and John Schleicher for their contributions on cooperative research projects. I also owe many thanks to all participants who devoted time to user studies in my research.

I would like to express special thanks to the Van Gogh Museum and the Kröller-Müller Museum for providing the datasets for my research.

I am also thankful for the College of Information Sciences and Technology and the US National Science Foundation for financially supporting my Ph.D. study.

Last, but by no means least, I thank my parents for their continued and unconditional support and encouragement. Without their dedication, I would not have travelled this far.
Dedication

To mama, who’s always there for me.
Chapter 1

Introduction

Visual data is the main subject studied in research communities like multimedia and computer vision. A core problem studied in these fields is whether computers can be endowed with the ability to perceive. The natural and straightforward task of seeing for humans has been demonstrated to be extremely difficult for machines. There exists a huge semantic gap between the raw data and the real world. Computational studies of the vision problem usually follow a paradigm as illustrated in Figure 1.1. Researchers aim to find solutions to bridge the semantic gap between the raw data and human perception. Approaches highly depend on the specific tasks and domain knowledge. Typical tasks include object recognition, content based image retrieval and annotation, motion analysis, and scene reconstruction. The ultimate goal is to emulate the cognition of human visual system and fulfil tasks based on successful perception. A major part of the research efforts are focused on the semantic analysis of images.

Besides semantics, aesthetics is also an essential aspect closely related to human perception and organization of visual data, especially for artistic images such as photographs and paintings. Aesthetics is originally a philosophy question concerning beauty and taste [8]. In the field of art and photography, aesthetics refers to the study of principles that underlie the appreciation of beauty. However, the study of aesthetics has not received much attention in the computer vision and image analysis research communities until recent years. Advances in technologies such as image acquisition, digital storage and computing have triggered the explosion of big visual data. Image data on the Internet has expanded at a tremendous
rate and continues to grow. Personal collections of digital images also expand due to accessibility to cheap storage and digital picture-taking devices. Visual information, which contains a great portion of the information we create and receive on a daily basis, has significantly impacted our lives. Semantic analysis has its limitation in visual data management and cannot meet the increasing user demands on image analysis tasks related to the appreciation of aesthetics. For example, traditional content based image retrieval algorithms focus on retrieving similar images as the query in terms of low level visual contents such as color and texture without considering aesthetics terms; the interpretation of digitized paintings is beyond the classic semantic analysis framework and requires understanding of principles in Art disciplines.

Traditionally, the appreciation of aesthetics in images is studied by professional artists, art historians, connoisseurs, photographers, designers and editors. Answers to questions like “what are the properties of an image where its aesthetic values rest?” and “how to create an image with high aesthetics quality?” have always been subjective attempts. To art professionals, the study and practice of aesthetics is usually time consuming and requires many manual efforts. When the task is contentious (for example painting authentication), the subjectiveness becomes an issue [28]. To the general public, the study and practice of aesthetics itself
is challenging. Therefore, studying computational approaches to aesthetics can potentially benefit both professionals and the general public. For example, art historians can expand their traditional toolbox with more efficient painting analysis software tools. Aesthetic assessment and enhancement methods for photos can assist photographers in photo collection management. Photo-centric websites can organize their large image databases in a more intelligent way and provide users with more intuitive experiences. In contrast to semantic analysis, the matter of subjects is usually deemphasized in the discussion.

Learning art and aesthetics is daunting, and computational studies of the richness and complexity of art can be intimidating. The major challenges in computational studies of aesthetics mainly result from: 1) the uncertainty and subjectiveness in measuring aesthetics value; 2) the huge gap between the data extracted and perceptive interpretation; 3) the existing difficulties of cross-disciplinary research; 4) the incomplete comprehension of the human visual system and related schemes of emotional responses. These fundamental problems are mentioned frequently in the literature. The semantic gap expresses the technical limitations of image understanding by computer. An analogous term aesthetics gap is defined in [19] as the lack of coincidence between the information that one can extract from low-level visual data and the aesthetic response or interpretation of emotions that the visual data may arouse in a particular user in a given situation. The variety in the appreciation of aesthetics exists across cultures, generations, individuals, and contexts. The perception of aesthetics depends on many different factors [14], too many factors and aspects to consider at one time. Therefore computational studies of aesthetics must focus on the deterministic elements identified by the particular research problem, and in most cases domain knowledge is particularly important. This dissertation mainly studies painting styles and photograph compositions.

The term composition, when applied to artistic creations such as writing, music, and painting, is usually interpreted as intentionally using basic elements to construct works of art. It contains at least two parts: the composing elements and the arranging of those elements, for example a poem formed by words and a piece of music created by notes. As for visual arts such as painting and photography, which are the main subjects studied in this dissertation, composing elements are visual, specifically line, color, shape, value, texture, and form as covered in many
books [58, 21, 25, 92] on visual arts. The principles of composition (or known as principles of design, principles of organization) involves almost all perspectives where considerations are given and choices are made by artists during the development of art works. An example list for principles of composition includes harmony, balance, unity, contrast, space, etc. In contrast to principles, techniques, or rules, usually refer to procedures that have been demonstrated to work effectively in practice. For example, the golden ratio and its simplified version “rule of thirds” are widely applied by artists and photographers. Other rules include “diagonal rule,” “simplicity,” “leading lines,” etc. These rules are effective and easy for amateurs to adopt, and can help amateurs in making well-balanced and attractive pictures. However we must note that these rules are not unalterable. To professional artists, strictly following the rules, on the contrary, limits the power of expression. Artists frequently and intentionally break rules to create more striking effects.

Composition is sometimes used as a synonym for artwork. For example, a good composition always refers to an aesthetically pleasing artwork. The composition of an image, as studied in this dissertation, is about the visual elements, principles, and techniques that are closely related to the aesthetic value of images. In recent years, the computational study of images has been extended to teach computers to appreciate the aesthetics of images. For example, models are computed from training datasets to guess the aesthetic value of any given photograph; rules and techniques are employed to modify or improve the model’s aesthetic quality. Although research in this domain is not very mature yet, recent achievements have shown much potential of developing real-world applications. Composition as a topic highly correlated to the aesthetics is therefore worthwhile to study.

While the composition of a work is the creative product of the artist’s thoughtful practice, “the notion of style has long been the art historian’s principal mode of classifying works of art. By style he selects and shapes the history of art”—Meyer Schapiros well-known account of style summarized by George Kubler and quoted in [44]. Style in visual arts, like painting, is often closely related to a period, a particular artist, an art school, or an art movement. Differentiation of styles are made across different career periods of an artist, artists, and art schools, which are the main subjects studied by art historians and connoisseurs in addressing problems such as attribution, authentication, and dating. Although scientific tools
such as chemical analysis, X-radiography, infrared technology, etc., are important, the interpretation of style, often a matter of subjective judgments by art historians and connoisseurs, remains crucial in painting analysis. As more paintings have been digitized, art historians are interested in exploring computational approaches as another scientific tool to characterizing artistic styles.

While we talk frequently about styles, we usually do so without serious thoughts of what is artistic style. What are the characteristics that distinguish Cubism, Expressionism, Impressionism, Pointillism, and Post-impressionism? What are van Gogh’s and Monet’s personal styles? “style is a characteristic or group of characteristics that we can identify as constant, recurring, or coherent. ... Artistic style is the sum of constant, recurring or coherent traits identified with a certain individual or group [31].” Style is the combination of many characteristics that distinguish career periods of an artist, different artists, and various art schools. This interpretation lays the basis for research on automated analysis of painting styles.

1.1 Contributions

Describing, extracting, and detecting visual elements are classic machine-vision problems. A tremendous number of works have been published in the literature that address this problem from different perspectives [18]. Based on the semantics, visual elements can be categorized as low level and high level. Low level elements are primitive features such as interest points, edges, regions, etc. The extraction of high-level elements and patterns are more interesting and challenging. This dissertation contributes to the study of aesthetics by focusing on the extraction and characterization of visual elements relevant to composition and style of paintings and photographs. Three specific problems are identified and studied.

First, automatic algorithms are proposed to extract and characterize the brush-strokes in oil paintings. I studied the painting styles of two artists, i.e. Vincent van Gogh and Norval Morrisseau. The algorithms are applied on real painting analysis tasks addressed by the art historians. Specifically, the problems studied include: 1) the identification problem—comparing the artist’s works with other contemporary artists (van Gogh); 2) the authentication problem—distinguishing
the artist’s authentic works from counterfeits (van Gogh and Norval Morrisseau); and 3) the dating problem—differentiating works developed in different career periods of a particular artist (van Gogh). In particular, algorithms are designed to extract the brushstrokes automatically from van Gogh’s paintings. Sets of visual features are computed to describe his brushstroke styles. Curve elegance measures are proposed to characterize the craftsmanship of the Canadian native artist Norval Morrisseau.

Secondly, algorithms are developed to classify general photographs into specific composition categories in terms of spatial layout. Specifically, algorithms are created to detect the diagonal compositional elements in photographs and identify horizontal/vertical/centered compositions. Rules in photography and visual perception are adopted in the design of algorithms. The classifier is further incorporated into photograph retrieval from databases to improve the relevance of returned photos. A photography feedback system is proposed to incorporate other photo analysis components in order to provide on-site feedback to amateur photographers.

Thirdly, an approach is proposed to emulate the dodging and burning techniques in photography. Inspired by concepts such as Notan in art and the darkroom dodging and burning processes, the proposed method attempts to improve the dark-light composition of digital photographs automatically by adjusting the dark-light proportion and intensity distribution in a region-wise fashion. An exemplar photograph is used to control the overall look of the modified image. The proposed approach can effectively transfer the illumination of the exemplar while considering aesthetic constraints.

1.2 Organization of this Dissertation

The dissertation is organized as follows: Chapter 2 reviews research in relevant topics. Chapter 3 and Chapter 4 focus on tackling two painting analysis tasks of two artists, specifically van Gogh (Chapter 3) and Norval Morrisseau (Chapter 4). Composition study of photographs is addressed in Chapter 5 and Chapter 6. Classification of photograph spatial layouts is introduced in Chapter 5. Digital dodging and burning are presented in Chapter 6.
A new discipline, computational aesthetics [35], has emerged in the last decade, gaining greater interest from the research community in the past few years. The essential questions that researchers want to ask are whether machines can appreciate/assess image aesthetics and, if so, is it possible to develop tools to assist in enhancing and even creating aesthetic values for images. The flooding of digital images on the Internet and the increase in personal collections create great demands for software tools to evaluate image aesthetics. It is difficult to articulate the aesthetic quality of images; even professional artists use abstract concepts or terms to describe images of high aesthetic quality, such as “good composition,” “well-balanced,” etc. Although aesthetics is a subjective concept, the population has great consensus on what is desirable, and that aesthetic consensus serves as both the basis and motivation for this research problem. In response, the key problem becomes the identification of image features related to human perception of aesthetics and then their mathematical representations. A definition of computational aesthetics is given in [35], which is quoted below:

Computational Aesthetics is the research of computational methods that can make applicable aesthetic decisions in a similar fashion as
humans can.

As an interdisciplinary field, computational aesthetics’ core is situated in computer vision and image processing. Ideas developed in the computing community are highly motivated by knowledge from other relevant disciplines such as art, photography, and psychology. Computational aesthetics has become an active research field in recent years. Promising results have been demonstrated to benefit professional and amateur artists, art researchers, and the general public. The rest of this chapter reviews relevant works in computational approaches to aesthetics studies in paintings and photographs.

2.1 Painting Analysis

Due to the advancement of image acquisition techniques, museums and other cultural institutions have digitized many artworks in their collections for the purposes of digital archiving, preservation, as well as providing accessibility of many precious artworks to more audiences, for example the Google Art Project [1] and Louvre website [2]. The digitalization of artworks also promotes the demand of automatic approaches to digital image analysis in the art domain, which leads to the active participation of researchers from scientific communities like image processing, computer vision, and computer graphics. Computational approaches can be employed and studied in the process of acquisition, achieving, and investigation. For example, image restoration and image retrieval for artworks, studying painting style or process, data acquisition and knowledge representation, authentication and dating problems, 3D scientific visualization and scene reconstruction, re-contextualization, etc. The main advantages of digital artwork analysis include not applying anything to the original work to safeguard for conservation and preservation, saving time and labor over using error-prone and time-consuming manual analysis, maintaining objectivity, leveraging the modern computational power, working on more than one paintings, etc. The traditional art connoisseurship has always been fraught with debates due to its subjectiveness. Computational analysis methods can potentially provide objective scientific evidences for painting analysis tasks. Although research on digital artwork analysis is still on its early stages, the results published in recent years are very promising.
A taxonomy of artistic image analysis, provided in [37], characterizes research from different abstract levels based on image content studied. Using the terms in that taxonomy, the work presented in this dissertation can be described as characterizing the geometric primitives “brushstroke” using shape and spatial arrangement descriptors which are then used to profile the “style” of a particular painting. To art historians, brushstrokes offer important visual clues and are seen as the signature of artists [39, 74, 86]. They are the most primitive structure that artists put on the canvas to construct sophisticated artistic forms. This section reviews previous works of automatic painting analysis with a focus on brushstroke analysis. Specific problems studied in this field include painting style characterization, image retrieval from digital painting databases, and reconstruction of original scenes when the artist was constructing the artwork. General surveys on painting analysis cover more topics [39, 79, 37].

The extraction of exact brushstrokes is very difficult due to aging, layer-on-layer, and blending painting techniques. The brushstrokes are often intersected by subsequent brush moves. Research on computational analysis of brushstrokes mainly falls into two categories: analysis based on profiling low-level features and analysis based on explicit brushstrokes extracted from images. The first approach characterizes the brushstroke patterns using low-level visual features. Techniques like shading and glazing used in painting suggests that texture features may be appropriate for brushstroke analysis in some painting genres [39]. Texture features derived from first-order statistics, co-occurrence matrix, and wavelet are used in [45] to distinguish brushstrokes drawn using dry materials from those drawn using fluid materials. 2-D multi-resolution Hidden Markov Models (MHMM) are used to characterize the spatial dependence among pixels [48] and the brushstroke style of a particular artist is profiled by a mixture of MHMM. Polatkan et al. [68] studied detection of forgeries on a dataset that contains pairs of original paintings and intentional copies by the same artist. Hidden Markov Tree modeling is applied on wavelet coefficients to capture the statistical structure of images [68]. In [53], digitalized paintings are decomposed into a pyramid of high/low frequency signals. High frequency band signals are predicted from neighboring signals of the same scale and adjacent scales using linear regression. The regression coefficients and prediction errors are used to characterize the drawing style of artists. Sparse coding
is adopted to describe the style of art works of artist Pieter Bruegel the Elder [36]. A set of basis functions is trained from a group of authentic works. Then a painting under examination is projected to these bases and kurtosis of the distribution of responses distinguishes the authenticity of the work. The more sparsely the basis can represent an unknown painting, the more likely it is authentic. Fractal analysis is employed as an authenticity tool in [83] to authenticate Jackson Pollock's poured works. The fractal composition of paintings is characterized by the fractal dimensions calculated by the box counting approach. Texton histograms are used to describe painting styles in [87]. Multiple texton codebooks are formed by the cluster centers of painting patches in a range of scales, and the style of a painting is profiled by texton histograms from different texton scales. One problem with low-level features is that they are prone to variations caused by the image acquisition processes. Besides, it is difficult to interpret these features intuitively.

The second approach is more intuitive and straightforward but limited to painting styles with visible and isolated brushstrokes. Although human eyes can identify brushstrokes with little effort, it is extremely difficult for computers to discern individual brushstrokes. Research on studying explicit brushstrokes are hitherto quite limited. Berezhony et al. applied a circular filter on van Gogh's paintings to enhance brushstroke contours [9]. Closed contours are then filled and thinned. The skeletons are approximated by high order polynomials, whose coefficients are taken as features of the brushstrokes. In a later work [10], the authors proposed to estimate the perceived prevailing orientation from the major axis of the contour-enhanced patches. The estimated brushstroke orientations are compared with human-perceived orientations for individual painting patches. A depicting approach is described in [75]. Pixels belonging to the top layer of brush strokes are identified, removed, and further impainted to reveal the next layer. The iteration continues until a smooth layer is reached. The entire process reveals the course of painting for the particular work. In [41], brushstrokes are derived from a plane graph constructed from the line-and-crossing responses yielded by difference of Gaussians (DoG) filters. A topological minor is obtained by dual graph contraction. Consecutive edges of the reduced graph are then grouped to form brushstrokes. Brushstroke analysis methods in this category characterize the painting style using descriptors derived from extracted geometric primitives. How-
ever, the existing approaches fail to extract explicit brushstrokes and therefore cannot formulate high-level brushstroke characteristics.

Currently the major focus of digital painting analysis lies in the study of approaches to computational depiction of painting styles and painting process rather than evaluating the aesthetic value of paintings. Domain knowledge plays a significant role, which inspires the designs of visual features and algorithms in most research. Importantly, the motivation to study computational approaches to painting analysis is to reinforce the toolbox of art historians and not to replace the traditional connoisseurship.

### 2.2 Aesthetics Assessment and Enhancement for Photographs

Photography started to be recognized as a fine art in 20th century. The aesthetics of photography was debated first because it produces images mechanically. Compared with paintings, photography is a rather young art discipline, but like painting and other visual arts, photography has developed into different schools. The master photographers founded theories and techniques from established principles of traditional art forms, optics, and long-term experiences. Due to the fundamental question about beauty, the identification of components that contribute to photographic aesthetic value is subjective. Either following the rules or breaking them is likely to create aesthetics. Photography as an art form reflects the reality in a more direct way; often the emotional chord struck with the audience is the resonance of the picture and memory of life. Therefore aesthetic quality that can be captured by computational methods is quite limited, especially when it is related to semantic understanding of the scene.

The analysis of user ratings on photo.net and dpcchallenge in [17] shows considerable agreement about photo aesthetics in the general public. The traceable techniques and principles in photography motivate ideas in computational studies of photo aesthetics. If we consider aesthetic quality of photographs from different aspects, namely technical skills (illumination, color, sharpness, etc), organization of low-level visual element (color, line, shapes, etc.), object-level composition
(sky, sea, human face, etc.), high-level semantics (event, emotion, etc.), then the consensus of aesthetic appeal is more likely to decrease with higher level factors incorporated. For example, most people may agree on a picture of some scenic landscape, but a group photo taken at a tourism site may only make sense to the acquaintance. The core problem here is to predict the aesthetic values based on the extractable visual information. It could be a binary classification between high-quality photos and low-quality photos, or finer-level evaluation of the aesthetic score. More intelligent photo-enhancement schemes are built upon the solution to the core problem.

2.2.1 Aesthetics Assessment

Much work has been done in the past to quantify the image quality based on low-level features especially in image compression and reconstruction [23, 91, 76, 101, 102]. Although evaluation of proposed metrics is usually conducted in comparison with subjective assessment, image quality mainly refers to the low-level qualities of images, such as sharpness, noise and color accuracy. These features are related to the aesthetic qualities of images but only from the perspective of technical skills. The problem of classifying photographs taken by professional photographers and home users is addressed in [84]. The low-level features they used can be grouped into four categories: color, energy, texture, and shape.

Assessment of higher-level aesthetics must also consider the emotional response an image will arouse in people. It was established in [17, 42] that photo aesthetics, despite being subjective, can be estimated using a set of images with a general learnt model which can predict the aesthetics of any image. The problem of aesthetic inference is upgraded from basic classification to score prediction. Ke, et al. propose a set of visual features to measure subjective assessment criteria for photographs including “simplicity,” “realism.” and “basic photographic technique,” and use Naive Bayes to predict the overall aesthetic quality for any given photo [42]. Datta et al. examined several photography rules including the rule of thirds, eye focus, depth of fields, shape convexity, and region composition [17]. The aesthetic quality of a photograph is assessed based on the principle of color harmony in [56]. Understanding aesthetics can aid many of the applications like summa-
rization of photo collections [60] and extraction of aesthetically pleasing images for image retrieval [59]. It can also be used as an intermediate step to steer the aesthetics enhancement process.

2.2.2 Aesthetics Enhancement by Cropping and Retargeting

Cropping is an important operation for photo enhancement, which improves photograph composition by adjusting subject location within the picture frame and removing undesirable content. Automatic cropping algorithms are developed to achieve these goals. Recommendation of a cropping window can be implemented by searching all possible cropping options which usually involves traversal of the parameter space, or by designing specific cropping strategies following some heuristic rules in art and photography. The search of the optimal cropping window is usually cast as an optimization problem. As in aesthetic assessment, visual features are usually motivated by photography rules or based on low-level visual features. Rule-driven approaches include consideration of faces [81, 103], composition templates [103, 80], rule of thirds [103, 12], and visual weight balance [12]. Low-level features, such as edge, texture, and color, have been used to build aesthetic models [57, 80]. Saliency-alike features have been used to preserve regions of interest [81, 103, 52]. In [16], composition rules are learned from the spatial distribution of concurrent patch pairs of arbitrary distance based on a massive professional photo dataset. The learned priors then serve as rules to guide the cropping of sub-views from wide-view photos.

Many other applications have been built by suggesting improvisations to the image composition through image retargeting. An approach to optimizing photograph composition by cropping and retargeting is proposed in [51]. Numerical measurements are defined to quantify how much a photograph conforms to photography principles such as rule of thirds, visual balance, diagonal dominance, etc. The algorithm optimizes a numerical score by searching the parameters of the crop window coordinates and the amount of inflation or deflation the image undergoes during the retargeting process. Saliency retargeting is used to enhance photo aesthetics in [95]. Luminance, color saturation, and sharpness are adjusted to modify
the saliency of different objects according to the saliency order specified by the user, and an aesthetics prediction model is employed to pick out the modification with the highest aesthetics score. An interactive application is presented in [13] that either recommends the locations for user-specified foreground objects so that a learned aesthetic metric is optimized or recommends sub-views or expansion for photographs without a distinct foreground object by equalizing the distribution of visual weights. A crop-and-warp-based scheme is proposed in [38] which locates a crop window by optimizing an energy function and then adopting the locations of salient objects and feature lines using the warping technique. Guo, et al. considered fidelity by minimizing changes made to image content as well as aesthetics by following the rule of thirds when optimizing image composition aesthetics [32]. Image warping method is also employed in [96] to relocate the subjects against the background to enhance the visual dominance of the main subjects.

Research on aesthetics assessment and enhancement generally follows the same paradigm as computer vision but focusing on an even higher-level perspective of human cognition. The current focus of study is mainly on investigating the correlation between aesthetics consensus existing in population and extractable features usually inspired by the principles in photography and relevant art disciplines. The visual features employed are low-level features, high-level features or a hybrid of the two. Models are built based on the features to emulate the human response to aesthetic qualities and suggest improvisation to image composition. The models are like black box and users have little control on the various contributing factors. The existing methods are mostly off-line in nature and therefore provide limited scope for improvement.

2.2.3 Exemplar-based Style Transfer

Another type of image enhancement is exemplar based approaches. As a particular type of style, color transfer studies the problem of applying the color palette of a target image to a source image, essentially reshaping the color distribution of the source image to accord with the target at some cost. The histogram matching algorithm derives a tone mapping function from the cumulative density functions (CDF) of the source and the target. The modified image then has a similar in-
tensity CDF as the target histogram. Minimizing the difference between the final output and the result given by histogram matching serves as an optimizing objective in [98] where minimizing the gradient difference is considered while preserving fidelity. Similarly both color differences and gradient differences are considered in [62] to achieve color transferring as well as preservation of the original look. The color transfer algorithm in [71] normalizes color distribution of the source image by the mean and the variance of the target image. Principal component analysis is utilized to model the transformation from the source color distribution to the target [5, 97]. Histogram transferring between two images is formulated as a mass transportation problem in [66]. An N-dimensional distribution transferring approach is proposed in [67], which performs color transfer by reshaping 1-D marginal distributions iteratively. In [69], minima and maxima of the source histogram and the target histogram are detected first, and then the mean and the variance of a target histogram region bounded by adjacent minima are used to reshape the corresponding histogram region in the source histogram. The number of different scales of histograms being reshaped controls the degree of color transfer.

In the broader sense, style transfer has been conceived in both photorealistic and non-photorealistic conditions. The framework of image analogy was first introduced in [34], where “styling” filters are trained from pairs of source images and styled images, and then applied to a different image to create a “styled” version. This method, however, requires a pair of well-registered images. A semi-supervised approach was proposed in [46]. Instead of using fully aligned image pairs, users only need to provide the algorithm with some styled patches as training data. Resales et al. proposed an unsupervised approach that casts the style transfer as a probability inference problem [72]. The most probable output image is estimated through belief propagation in Markov random field. In [89], photo pairs, consisting of a snapshot taken by a low-cost camera and its counterpart taken by a high-quality camera with the same subject, are used to train a regression model that transfers the high quality style to photos taken by low-cost cameras. A photorealistic tone management was presented in [7] where an image is decomposed into a base layer and a detail layer; the two layers are then manipulated separately for style transfer.

The methods identified above process the color distribution globally and do
not consider spatial information. Pixels of the same intensity level are subjected to the same transformation regardless of whether they are in dark regions or light regions. Artifacts can be easily brought in when the source histogram is very different from the target. Color transfer is then conducted between corresponding regions in the source image and the target image [82, 93], however, these methods either need manual inputs or require much computation. In [93], user drawn strokes are used to specify corresponding regions. Tai et al. generate probabilistic segmentation in images by a modified expectation maximization algorithm, and the correspondences between regions in the source image and the target is dened according to luminance order for general cases and spatial overlap for content-wise similar images [82].
Distinguish van Gogh from his contemporaries via Automated Brushstroke Extraction

3.1 Introduction

Art historians utilize a variety of tools and methods in their studies such as chemical analysis, X-radiation, infrared technology, documentary research, painting style analysis, etc. The interpretation of styles shows the inner characteristics of artworks and makes critical judgement in painting analysis. As more paintings have been digitized, art historians are interested in exploring computational approaches as another scientific tool to characterizing artistic styles. One early paradigm in connoisseurship by Giovanni Morelli is studying individual attributes of artworks by the same artist and identifying particular forms by the artist. For example, if every aspect of a painting under examination resembles the characteristics observed in other authentic paintings, then there is less doubt about the authentication of the particular artwork. Computational approaches to painting style characterization usually follows this paradigm. Many efforts have been made by the computing community to provide art historians with scientific evidences potentially valuable in assisting traditional painting analysis tasks including artist identification, dat-

The materials in this chapter have appeared in [50].
Art historians and connoisseurs consider brushstrokes as artistic signatures. The application of rhythmic and spontaneous brushstrokes is a prominent characteristic of Van Gogh’s paintings [40] that art historians utilize in authentication of his artworks. Recent works on painting analysis have investigated feature extraction methods focused on characterizing brushstrokes [41, 10, 75], but extraction of brushstrokes for paintings in general has not been well studied. Most studies focus on profiling low-level features but fail to depict high-level brushstroke characteristics. This chapter presents an automatic brushstroke extraction scheme based on edge detection and clustering-based image segmentation, and propose a group of brushstroke features. The brushstroke extraction algorithm and brushstroke feature are applied on analysis of van Gogh’s paintings.

Two challenges designed by art historians are studied in this work. The first challenge is to separate van Gogh from his contemporaries (Figure 3.1) by comparing characteristics of van Gogh’s brushworks with other artists. Paintings by other artists were not deliberate copies or forgeries but have been mistakenly attributed to van Gogh for some reason. For this challenge, art historians suggested two groups of paintings (four in each group) for the comparative study.

Figure 3.1: van Gogh’s paintings versus his contemporaries’.
- Paintings by van Gogh: Portrait of a Young Girl Against a Pink Background (painting ID F518, Auvers, late June-early July 1890), Chestnut Tree in Flower: White Blossoms (F752, Auvers, May 1890), Still Life: Vase with Rose Mallows (F764a, Auvers, June 1890), View at Auvers (F799, May-June 1890).

- Paintings by van Goghs contemporaries: Red Cliffs near Anthéor (S447, by Louis Valtat, c. 1903), Schönbrunn (S448, by Carl Moll, c. 1910), Garden with Hollyhock (S457, by Ernest Quost, before 1888), and Mills at Westzijderwald near Zaandam (S503, by Claude Monet, 1871).

The painting identifications are based on the catalogue numbers in the revised edition of the oeuvre catalogue by J.-B. de la Faille [20] for van Gogh’s paintings, and the inventory numbers of the Van Gogh Museum collection for his contemporaries’.

Figure 3.2: van Gogh’s painting from Paris period.

The second challenge addressed by art historians was to date a set of van Goghs paintings into two periods of his development: Paris Period (Figure 3.2) vs. Arles and Saint-Rémy Period (Figure 3.3). Eight paintings from each of the two periods are studied. Three paintings under debate (Figure 3.4) are investigated in particular.

- Paris works includes eight paintings all dated to 1887 [11, 33]: A Skull (F297, May-June 1887), Still Life: Romans Parisiens (F358, October-November
Figure 3.3: van Gogh’s painting from Arles-Saint-Rémy period.

Figure 3.4: Three paintings that receive different opinions on dating.

1887), Still Life with Plaster Satuette, a Rose and Two Novels (F360, late 1887), Japonaiserie: The Flowering Plum Tree: after Hiroshige (F371, October-November 1887), Red Cabbage and Onions (F374, November 1887-February 1888), Four Cut Sunflowers (F452, August-October 1887), Self-Portrait with Straw Hat (F469, August-September 1887), and Self-Portrait with Pipe and Straw Hat (F524, September-October 1887).

- Seven paintings dated to 1888 in Arles [88, 11, 22] and one dated to late 1889 in Saint-Rémy [4]: Blossoming Almond Branch in a Glass (F392, March 1888), Wheatfield (F411, June 1888), Seascape at Saintes-Maries (F415, June 1888), The Baby Marcelle Roulin (F441, December 1888), The Sower (F451, c. 25 November 1888), The Green Vineyard (F475, c. 3 October 1888), Portrait of Camille Roulin (F538, December 1888), and Leather Clogs (F607,
late 1889).

- Three paintings that receive different opinions on dating: Still Life: Potatoes in a Yellow Dish F386 (formerly dated to the Paris period but recently considered one of the earliest works in Arles), Willows at Sunset F572 (similar debate as F386), and Crab on its Back F605 (the current dating to January 1889 was questioned because of R. Pickvances assertion that the related picture of Two Crabs should be relocated from January 1889 to the late Paris period [65]).

The digitized paintings used in this study are provided by the Van Gogh Museum and the Kröller-Müller Museum. Color large-format transparency films of the original paintings were scanned at high resolution and scaled to a uniform density of 196.3 dots per painted-inch and digitized to 16 bits per channel. Due to copyright issues, the right half of each digitized painting was checkerboarded.

### 3.2 Brushstroke Extraction

The brushstroke extraction process is divided into two passes: one based on edge extraction and the other based on image segmentation. Observable brushstroke boundaries are likely to be caught by edge detection algorithms. However, it is extremely difficult to identify the complete outline of a brushstroke, because a brushstroke can be overlaid by other brushstrokes, the boundaries of adjacent brushstrokes may be blended together, ridges formed within a brushstroke due to the painting techniques, etc. Often the edges obtained by edge detection algorithms are only aligned with part of the complete brushstroke outlines. The edges may be broken by gaps of undetected boundaries or mistakenly merged with edges nearby. In order to recover those partially detected brushstrokes, morphological operations are employed to form closed contours for nearly finished outlines. Although this approach is intuitive and robust, brushstrokes whose boundaries are not sharp enough to be caught by the edge detection algorithm will still be left out. A subsequent pass based on image segmentation is used to gain more brushstrokes from the rest of the image.
Figure 3.5: Flowchart of the brushstroke extraction algorithm.

3.2.1 Algorithm Overview

Figure 3.5 illustrates the flowchart of the brushstroke extraction procedure. The brushstrokes extracted from the two separate passes are combined at the end for further analysis. The algorithms for each component will be described later in more detail.

3.2.1.1 Edge Detection Based Brushstroke Extraction

A confidence based edge detection algorithm by Meer and Georgescu [55] is applied on images to identify edge pixels. A gradient and a confidence measure are computed for each pixel, and the nonmaxima suppression and hysteresis thresholding steps in the edge detection paradigm are conducted in the gradient-confidence plane. The parameters that need to be specified include window size for gradient
Figure 3.6: Brushstroke extraction based on edge detection. (a) A painting patch. (b) Edge pixels detected by the edge detection algorithm in [55]. (c) After short edges (less than 30 pixels) are removed. (d) After enclosing small edge gaps (distance less than 15).

computation, the minimum edge length, parameters for nonmaxima supression and hysteresis thresholding. The window size and minimum edge length are set to be 3 and 50 respectively for all images. The parameters for nonmaxima suppression and hysteresis thresholding correspond to three curves in the gradient-confidence plane. Elliptical curves is chose as the form of thresholding curves. The thresholds (including gradient and confidence) in nonmaxima supression are set to be 0.1. The lower bounds (including gradient and confidence) and higher bounds in hysteresis thresholding are set to be 0.1 and 0.2 respectively. After edge detection, an edge linking algorithm [43] is used to trace down edge pixels. In this way, pixels that belong to the same edge are linked together in the tracing order. Short edges are removed after the linking stage. Figure 3.6 shows an example of using the edge detection [55] and edge linking [43] to acquire valid edges.

The edges corresponding to the brushstroke contour are usually not completely detected (Figure 3.6 (c)). An enclosing operation is designed to complete the nearly closed gaps between edges. The endpoint of any detected edge is examined to see if there exist other edge pixels in its neighborhood which do not belong to the edge under consideration. If such edge pixels are found in the neighborhood, the endpoint is then connected to the closest one by a straight line. The isolated connected components formed by non-edge pixels enclosed by edges are considered as brushstroke candidates. Figure 3.6 (d) shows the enclosed edge map for the given painting patch. Pixels in a particular color belong to the same connected
3.2.1.2 Image Segmentation Based Brushstroke Extraction

Images of paintings are divided into color homogeneous regions by the clustering based image segmentation algorithm in [47]. The 5 dimensional input feature vector for the segmentation algorithm includes three color channels in RGB color space and intensity gradients in both horizontal and vertical directions. The segmentation algorithm generates clusters of pixels by multiple iterations of k-means with decreasing thresholds for within-cluster distance. Figure 3.7 presents segmentation results for the same painting patch shown in Figure 3.6. Pixels in extracted brushstrokes (obtained in the first pass using the edge detection based approach) are no longer considered in the second pass. Connected components are extracted from the remaining pixels.

3.2.1.3 Brushstroke Test

The connected components obtained by the previously described methods, i.e. edge detection based approach and image segmentation based approach, are subjected to a brushstroke test. The test includes the following criteria. Connected components that satisfy all the criteria are considered as valid brushstrokes.

1. The size of the connected component should be within a preselected range which is set to be [100, 800] in this work. The intuition behind is that brushstrokes should have a reasonable coverage. Segments which are too large

Figure 3.7: Brushstroke extraction based on image segmentation. Left: image segmentation results. Right: pixels that belong to brushstrokes extracted in the first pass removed.
or too small are more likely to be results of under-segmentation or over-segmentation. The employed range should be adapted to the actual image resolution.

2. The skeleton of the connected component should not be severely branched. Based on observation, a valid brushstroke is usually in the shape of an elongated strip. A connected component with many bumps or indentations is more likely to be the space between brushstrokes or results of under-segmentation or over-segmentation. A brushstroke with one dominant backbone is considered as not severely branched. The backbone extraction algorithm will be described later.

3. The shape of a valid brushstroke is usually an elongated strip, therefore the ratio between broadness and length is thresholded to filter out squarelike or circular connected components. A range of $[0.05, 1.0]$ for the ratio is used for this purpose.

4. For a regularly shaped brushstroke, the size of the connected component is approximately length multiplied by width. The last criterion is to check whether the ratio between the size of the connected component and the value of brushstroke length (length of the backbone) multiplied by the width (twice the maximum distance from boundary pixels to the backbone) is within a range (set to be $[0.5, 2]$ in this work).

### 3.2.1.4 Backbone Extraction

The skeleton can be considered as the essence of the connected component which characterizes its shape and structure. A skeleton is obtained by a morphological thinning algorithm. Most of the skeletons extracted from the connected components contain unwanted branchy components. A skeleton is considered not severely branched if one single dominant backbone exists, i.e. the branches are comparatively much smaller than the main backbone. The edge linking algorithm [43] is used here to trace down the skeleton and record the skeleton pixels in order. For a skeleton without any branches, the edge linking algorithm discerns the backbone by tracing down the skeleton from one endpoint to the other. On the other hand,
Figure 3.8: Backbone extraction. (a) A branched skeleton. (b) Two edge segments after edge linking. (c) Two artificial endpoints (in red). (d) Edge segments after inserting artificial endpoints. (e) Backbone.

for a branched skeleton, multiple edge segments will be generated by the edge linking algorithm without identification of the backbone. Figure 3.8 illustrates the process of discovering the backbone from a branched skeleton. Artificial endpoints are inserted at the branching point of the skeleton (Figure 3.8 (c)). The shortest edge segment is removed and the two longer edge segments are merged into a backbone. The algorithm of backbone extraction in general has two main steps.

1. Insert artificial endpoints at the branching positions. Pixels on edge segments of a skeleton are examined in the tracing order resulted from edge linking. A pixel on the skeleton becomes an artificial endpoint if it satisfies three conditions: (1) the pixel is not an endpoint; (2) the pixel is not a 8-connected neighbor of either end point of the edge segment it belongs to; (3) the pixel is a 8-connected neighbor of an endpoint of another edge segment. If a pixel meets the three conditions, it becomes a new endpoint and the pixel next to it in tracing order is also a new endpoint. The newly inserted endpoints then break the edge segment into two new edge segments (Figure 3.8 (d)).

2. Merge edge segments at branching positions. An endpoint is marked as a branching position if it has two or more neighbors which are endpoints of different edge segments. At the branching position, the two longest edge segments are merged while short ones are removed from the skeleton. Then there are no more endpoints at this branching position because they are either removed or become intermediate points in the merged edge segment.
Figure 3.9: Examples for extracting backbones in branched skeletons. First row: an example for “not severely branched” skeletons. Second row: an example for a “severely branched” skeleton. (a) Connected components. (b) The skeletons shown as edge segments formed by edge linking. (c) Edge segments formed after inserting artificial endpoints. (d) Edge segments left after merging at branching positions.

This process continues until no more branching positions can be found.

After processing all the branching positions, if two or more edge segments are left, this skeleton is considered to be severely branched. Otherwise, it is not severely branched. “Not severely branched” is only one criterion for the brushstroke test. A connected component has to satisfy all criteria enumerated in the previous section to be claimed a valid brushstroke. Figure 3.9 illustrates the backbone extraction process for a “not severely branched” case and a “severely branched” case. The two examples are connected components from real paintings.

3.2.2 Evaluation of Brushstroke Extraction

The brushstroke results obtained for some of the full paintings provided by the Van Gogh Museum are shown in Figure 3.10.

In order to evaluate the accuracy of the brushstroke extraction algorithm, automatically extracted brushstrokes are compared with manually marked brush-
Figure 3.10: Brushstroke extraction results for van Gogh paintings. Painting images courtesy of the Van Gogh Museum Amsterdam (Vincent van Gogh Foundation).
strokes. It requires considerable efforts to manually delineate individual brushstrokes because one has to trace the brushstroke boundary by hand and a painting can easily contain thousands of brushstrokes. In the comparative study, 10 patches cropped from paintings in the Van Gogh collection are used as test data. The observable brushstrokes in the 10 patches are manually marked out by a student. The sizes of the patches range from $300 \times 171$ to $300 \times 274$, and an average of 120 brushstrokes are manually marked for each patch. The brushstroke test described previously is applied on the manually marked brushstrokes, and the average pass rate is 95%, which justifies the criteria in the brushstroke test. The pass rates for individual painting patches are reported in Table 3.1.

Direct comparison is also made between the automatically extracted brushstrokes and the manual brushstrokes. For a painting patch in consideration, suppose $n$ brushstrokes, $B_i$, $i = 1, \ldots, n$, are found by the extraction algorithm, and $m$ brushstrokes, $B'_j$, $j = 1, \ldots, m$, are marked manually. $B_i$ ($B'_j$) is the set of pixels in the brushstroke. $B_i \cap B'_j$ is the set of overlapped pixels in the two brushstrokes. Then $B_i$ is claimed being validly covered (or for brevity, covered in the sequel) by $B'_j$ if the overlap between the two makes up more than 80% of pixels in $B_i$, that is, $|B_i \cap B'_j|/|B_i| > 80\%$. Let $C_{i,j} = 1$ indicate that $B_i$ is covered by $B'_j$, and $C_{i,j} = 0$ otherwise. Obviously, $B_i$ can be covered by at most one manual brushstroke. Define $C_{i,\cdot} = \sum_{j=1}^{m} C_{i,j}$, which indicates whether $B_i$ is covered by any manual brushstroke at all. $C_{i,\cdot} \in \{0,1\}$. If $C_{i,\cdot} = 1$, then $B_i$ is valid. Define $C_{\cdot,j} = \sum_{i=1}^{n} C_{i,j}$, which is the number of automatically extracted brushstrokes that are covered by manual brushstroke $B'_j$. $C_{\cdot,j} \in \{0,1,\ldots,n\}$. $B'_j$ is claimed being detected if $C_{\cdot,j} \geq 1$. The accuracy of automatic brushstroke detection is measured by two quantities: valid rate $r_v$—the percentage of valid automatically extracted brushstrokes and detection rate $r_d$—the percentage of detected manual brushstrokes. Both quantities are calculated by equations in (3.1). The valid rate and the detection rate of the 10 painting patches are reported in Table 3.1.

$$r_v = \frac{\sum_{i=1}^{n} C_{i,\cdot}}{n}$$
$$r_d = \frac{\sum_{j=1}^{m} I(C_{\cdot,j} \geq 1)}{m}$$  \hspace{1cm} (3.1)
Table 3.1: Comparing Automatically Extracted and Manually Marked Brushstrokes in 10 Patches Cropped from Paintings in the van Gogh Collection.

<table>
<thead>
<tr>
<th>Painting ID</th>
<th>Pass Rate</th>
<th>Valid Rate</th>
<th>Detection Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>F218</td>
<td>97.4</td>
<td>42.7</td>
<td>21.6</td>
</tr>
<tr>
<td>F248a</td>
<td>83.6</td>
<td>75.4</td>
<td>78.8</td>
</tr>
<tr>
<td>F297</td>
<td>95.6</td>
<td>57.9</td>
<td>52.0</td>
</tr>
<tr>
<td>F374</td>
<td>96.0</td>
<td>58.2</td>
<td>63.2</td>
</tr>
<tr>
<td>F386</td>
<td>97.8</td>
<td>73.7</td>
<td>68.4</td>
</tr>
<tr>
<td>F415</td>
<td>90.9</td>
<td>46.9</td>
<td>60.0</td>
</tr>
<tr>
<td>F518</td>
<td>97.2</td>
<td>60.7</td>
<td>75.2</td>
</tr>
<tr>
<td>F538</td>
<td>98.0</td>
<td>49.0</td>
<td>44.9</td>
</tr>
<tr>
<td>F572</td>
<td>96.6</td>
<td>83.9</td>
<td>65.6</td>
</tr>
<tr>
<td>F652</td>
<td>95.9</td>
<td>50.0</td>
<td>72.5</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>94.9</strong></td>
<td><strong>59.8</strong></td>
<td><strong>60.2</strong></td>
</tr>
</tbody>
</table>

Figure 3.11 shows the 10 patches cropped from paintings by van Gogh. The automatically extracted brushstrokes and the manually marked brushstrokes are displayed on the side. In these examples, the automatically extracted brushstrokes may not correspond precisely to the physical brushstrokes. Sometimes, the ridges and the under paint from one manually marked brushstroke may be extracted as multiple brushstrokes. Sometimes, only a portion of a manually marked brushstrokes is extracted automatically. Despite the disparity with physical brushstrokes, the brushstrokes extracted by the algorithm appear to capture well the characteristics of the textured patterns created by the paint, for example, orientation and richness of color. Indeed, it is difficult even for art historians to decipher the painting process down to a brushstroke level for a medium as complex as oil paint.

### 3.3 Brushstroke Features

The output of the brushstroke extraction algorithm is the connected components that pass the brushstroke test. Those connected components are claimed valid brushstrokes (brushstroke for brevity in the sequel). Numerical features are designed to characterize brushstrokes. Motivated by expert opinions from art histori-
Figure 3.11: Comparison of automatically extracted brushstrokes and manually marked brushstrokes. (a) Painting patch. (b) Automatically extracted brushstrokes. (c) Manually marked brushstrokes. Painting IDs from top to bottom: F218, F248a, F297, F374, F386.
Figure 3.11 cont. Painting IDs from top to bottom: F415, F518, F538, F572, F652.
After the extraction of brushstrokes, numerical features can be computed. The skeletons and boundaries of the brushstrokes are shown.

Two types of features are calculated for each brushstroke: geometric features and interactive features. The geometric features characterize the shape of the brushstroke while interactive features describe how it relates to its neighboring brushstrokes. There are in total seven geometric features: length, width (or broadness), size, broadness homogeneity, elongatedness, straightness, and orientation; and three interactive features including the number of brushstrokes in the neighborhood, the number of brushstrokes with similar orientations in the neighborhood, and the amount of variation (measured by standard deviation) in the orientations of brushstrokes in the neighborhood. The size of the neighborhood is fixed for all paintings, i.e. $401 \times 401$ centered at the centroid of the brushstroke.

The brushstroke features are calculated based on pixels within the connected component. Two structural elements are used in the calculation: the boundary and the backbone (Figure 3.12). The backbone of a brushstroke has been obtained using the thinning and backbone extraction operation. The coordinate of a pixel in the digitized painting is $(u, v)$ where $u = 0, 1, ..., R - 1$ and $v = 0, 1, ..., C - 1$. $R$ and $C$ are the total number of rows and the total number of columns in the image respectively. Denote the centroid of $i$th brushstroke by $(\bar{u}_i, \bar{v}_i)$, where $\bar{u}_i$ is the
average vertical position of all the pixels in the brushstroke and \( \bar{v}_i \) is the average horizontal position. The brushstroke features are:

- **Number of Brushstrokes in the Neighborhood (NBS-NB):** The \( j \)th brushstroke is a neighbor of the \( i \)th brushstroke if \(|\bar{u}_i - \bar{u}_j| < s \) and \(|\bar{v}_i - \bar{v}_j| < s\), where \( s \) is a threshold set to 200 in the experiments. NBS-NB for the \( i \)th brushstroke is obtained by counting the number of its neighboring brushstrokes.

- **Number of Brushstrokes with Similar Orientations in the Neighborhood (NBS-SO):** Two brushstrokes are considered to have similar orientations if the difference between their orientations is below a threshold, set to 0.35 in the experiments.

- **Orientation standard deviation for brushstrokes in a neighborhood (OSD-NB):** For the \( i \)th brushstroke, compute the standard deviation for the orientations of the brushstrokes in its neighborhood.

- **Size:** The size of a brushstroke is the number of pixels in the brushstroke.

- **Length:** The length of a brushstroke is the number of pixels on the backbone of the brushstroke.

- **Broadness:** The broadness of a brushstroke is the average Euclidean distance on the image plane from all boundary pixels to the backbone of the brushstroke. The distance between a boundary pixel and the backbone is the minimum distance between the boundary pixel and any pixel on the backbone.

- **Broadness Homogeneity (BH):** For every boundary pixel in the brushstroke, find its distance to the backbone of the brushstroke. The standard deviation of these distances normalized by the broadness of the brushstroke is defined as BH. The smaller the value, the greater the homogeneity.

- **Straightness:** The straightness of the brushstroke is computed as the absolute value of the linear correlation coefficient between the horizontal and vertical coordinates of the pixels located on the backbone of the brushstroke. The
correlation coefficient of a perfect straight line has an absolute value of one. The absolute value of the coefficient of a curved line will be smaller than one. Suppose a brushstroke contains \( N \) pixels with coordinates \((u_i, v_i)\), \( i = 1, \ldots, N \). The straightness is defined by \( |S_{uv}|/(S_u \cdot S_v) \), where

\[
S_{uv} = N \sum_{i=1}^{N} u_i v_i - \sum_{i=1}^{N} u_i \sum_{i=1}^{N} v_i,
\]

\[
S_u = \sqrt{N \sum_{i=1}^{N} u_i^2 - \left( \sum_{i=1}^{N} u_i \right)^2},
\]

\[
S_v = \sqrt{N \sum_{i=1}^{N} v_i^2 - \left( \sum_{i=1}^{N} v_i \right)^2}.
\]

- Elongatedness: The measure for elongatedness is defined as the ratio between the length and the broadness.

- Orientation: The definition given by J. C. Russ \([73]\) is used. The orientation of an area is essentially that of its principal axis. Let \( m_u = \sum_{i=1}^{N} u_i^2 - \frac{1}{N} \left( \sum_{i=1}^{N} u_i \right)^2 \), \( m_v = \sum_{i=1}^{N} v_i^2 - \frac{1}{N} \left( \sum_{i=1}^{N} v_i \right)^2 \), \( m_{uv} = \sum_{i=1}^{N} u_i v_i - \frac{1}{N} \sum_{i=1}^{N} u_i \sum_{i=1}^{N} v_i \). The orientation is computed as follows:

\[
\begin{cases}
\frac{\pi}{2} & \text{if } m_{uv} = 0; \\
\arctan \frac{m_u - m_v + \sqrt{(m_u - m_v)^2 + 4m_{uv}^2}}{2m_{uv}} & \text{otherwise}. 
\end{cases}
\]

The features from the list above describe brushstrokes at the individual brushstroke level. A painting may contain thousands of brushstrokes. To compare the brushstroke styles at the painting level, some statistics are derived from the features of all brushstrokes in one painting. It is observed that van Gogh’s brushstrokes do not have similar orientations across paintings. Therefore the amount of variation in orientation, measured as the standard deviation of the orientation across all brushstrokes in a painting, instead of the average orientation is used as the painting-level attribute, which is considered to be related to whether a painting conveys a unified look. For other features, the average is obtained as the painting-level attribute. The average brushstroke length and broadness is normalized by the
square root of the painting size (width $\times$ height); while the average brushstroke size is normalized by the painting size. The total number of brushstrokes in a painting is also included as a painting-level attribute. Therefore, there are in total eleven attributes computed for each painting: Total Number of Brushstrokes (TNBS), Number of Brushstrokes in the Neighborhood (NBS-NB), Number of Brushstrokes with Similar Orientations in the Neighborhood (NBS-SO), Orientation Standard Deviation in the Neighborhood (OSD-NB), Broadness Homogeneity (BH), elongatedness, straightness, length, broadness, size, and Orientation Standard Deviation (OSD).

### 3.4 Statistical Analysis and Findings

Art historians are interested in identifying attributes that can distinguish paintings in two groups, namely van Gogh (referred to as vG) vs. his contemporaries (referred to as non-vG) in the first study and Paris period (referred to as vG-Paris) vs. Arles-Saint-Rémy period (referred to as vG-Arles) in the second. Let the null hypothesis be that the two groups have the same average for a particular attribute. A two-sided permutation test is employed to compute p-values for the two comparative studies described in the previous section. In a two-sided permutation test, the paintings in two groups are shuffled to form two random groups. Let $x_1, x_2, ..., x_n$ and $x_{n+1}, x_{n+2}, ..., x_{n+m}$ be the attribute values for paintings in the first group and the second group respectively. After a random shuffle of $x_1, x_2, ..., x_{n+m}$, the shuffled attribute values form two random groups $x^{(1)}_1, x^{(2)}_2, ..., x^{(n)}_n$ and $x^{(n+1)}_{n+1}, x^{(n+2)}_{n+2}, ..., x^{(n+m)}_{n+m}$. The difference between the average attribute values of the two groups is $\delta_o$ for the two original groups and $\delta_p$ for the two groups after random permutation. Repeat the random permutation. The proportion of random permutations that result in $\delta_o < \delta_p$ is the p-value for the permutation test. The smaller the p-value is, the stronger the evidence is against the null hypothesis. The p-values for the eleven attributes under each of two studies are reported in Table 3.2 and 3.3 respectively. The statistical analysis in [47] leads to the following conclusions regarding the two problems addressed by art historians.

- The marking attributes of van Gogh vs. non van Gogh do not overlap with
those distinguishing van Goghs Paris and Arles/Saint-Rémy periods.

- The four marking attributes of van Gogh vs. non van Gogh are: NBS-NB, elongatedness, straightness, and BH.

- The five marking attributes of Paris vs. Arles/Saint-Rémy are: length, size, broadness, OSD-NB, OSD. The last two attributes result from subject-wise difference in the two periods, while the first three are caused by styles.

- Although the copy of Two Children (S506, by Cuno Amiet, 1907) is similar to van Gogh in terms of individual brushstroke attributes, elongatedness and straightness, it is closer to non van Gogh in terms of the interactive brushstroke attribute NBS-SO.

- The paintings F386 (Still Life: Potatoes in a Yellow Dish) and F605 (Crab on its Back) are dated to the Arles and Saint-Rémy period, while the painting F572 (Willows at Sunset) is dated to the Paris period.

Table 3.2: P-values for the eleven attributes under the study that compares paintings by van Gogh and his contemporaries.

<table>
<thead>
<tr>
<th>Painting-level Attributes</th>
<th>van Gogh vs. non van Gogh</th>
<th>VG landscape vs. non van Gogh</th>
<th>All VG vs. Non-VG</th>
</tr>
</thead>
<tbody>
<tr>
<td>TNBS</td>
<td>0.457</td>
<td>0.086</td>
<td>0.030</td>
</tr>
<tr>
<td>NBS — NB</td>
<td><strong>0.029</strong></td>
<td>0.029</td>
<td><strong>0.002</strong></td>
</tr>
<tr>
<td>NBS — SO</td>
<td>0.114</td>
<td>0.171</td>
<td><strong>0.000</strong></td>
</tr>
<tr>
<td>OSD — NB</td>
<td>0.114</td>
<td>0.543</td>
<td><strong>0.002</strong></td>
</tr>
<tr>
<td>elongatedness</td>
<td><strong>0.029</strong></td>
<td>0.029</td>
<td><strong>0.000</strong></td>
</tr>
<tr>
<td>straightness</td>
<td><strong>0.029</strong></td>
<td>0.029</td>
<td><strong>0.000</strong></td>
</tr>
<tr>
<td>BH</td>
<td>0.057</td>
<td>0.057</td>
<td><strong>0.000</strong></td>
</tr>
<tr>
<td>length</td>
<td>0.857</td>
<td>0.343</td>
<td>0.668</td>
</tr>
<tr>
<td>size</td>
<td>0.571</td>
<td>0.257</td>
<td>0.661</td>
</tr>
<tr>
<td>broadness</td>
<td>0.343</td>
<td>0.171</td>
<td>0.122</td>
</tr>
<tr>
<td>OSD</td>
<td>0.114</td>
<td>0.314</td>
<td><strong>0.006</strong></td>
</tr>
</tbody>
</table>
Table 3.3: P-values for the eleven attributes under the study that compares van Gogh’s paintings in Paris period and in Arles-Saint-Rémy period.

<table>
<thead>
<tr>
<th>Painting-level Attributes</th>
<th>Paris vs. Arles-St. Rémy</th>
<th>Paris vs. Arles-St. Rémy landscape vs. still life and portrait</th>
<th>Arles-St. Rémy landscape vs. still life and portrait</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>still life and portrait</td>
<td></td>
</tr>
<tr>
<td><strong>TNBS</strong></td>
<td>0.561</td>
<td><strong>0.059</strong></td>
<td>0.200</td>
</tr>
<tr>
<td><strong>NBS – NB</strong></td>
<td>0.852</td>
<td>0.537</td>
<td>0.343</td>
</tr>
<tr>
<td><strong>NBS – SO</strong></td>
<td>0.449</td>
<td>0.566</td>
<td>1.000</td>
</tr>
<tr>
<td><strong>OSD – NB</strong></td>
<td><strong>0.069</strong></td>
<td>0.588</td>
<td><strong>0.057</strong></td>
</tr>
<tr>
<td><strong>elongatedness</strong></td>
<td>0.208</td>
<td><strong>0.018</strong></td>
<td><strong>0.086</strong></td>
</tr>
<tr>
<td><strong>straightness</strong></td>
<td>0.837</td>
<td>0.921</td>
<td>0.743</td>
</tr>
<tr>
<td><strong>BH</strong></td>
<td>0.763</td>
<td>0.671</td>
<td>0.229</td>
</tr>
<tr>
<td><strong>length</strong></td>
<td><strong>0.097</strong></td>
<td><strong>0.014</strong></td>
<td>0.114</td>
</tr>
<tr>
<td><strong>size</strong></td>
<td><strong>0.090</strong></td>
<td><strong>0.014</strong></td>
<td>0.143</td>
</tr>
<tr>
<td><strong>broadness</strong></td>
<td><strong>0.080</strong></td>
<td><strong>0.016</strong></td>
<td>0.200</td>
</tr>
<tr>
<td><strong>OSD</strong></td>
<td><strong>0.062</strong></td>
<td>0.576</td>
<td><strong>0.086</strong></td>
</tr>
</tbody>
</table>

### 3.5 Summary

In this chapter, a novel brushstroke extraction method was described which exploits an integration of edge detection and clustering-based segmentation. The proposed algorithm has been used to automatically extract brushstrokes from van Gogh’s paintings. In this work, van Gogh’s painting style is compared with his contemporaries by statistically analyzing a massive set of extracted brushstrokes. The results clearly suggest that rhythmic brushstrokes in van Gogh’s paintings distinguish his work from those of his contemporaries, which aligns with long-held art historical opinion on van Gogh’s unique style of painting. For the first time though, the information is presented in a quantitative way, providing more refined and accurate data to substantiate the art historians’ opinion. Furthermore, these new techniques were applied to compare brushwork characteristics in three paintings that have proved difficult to date by art historians using traditional means. This provided new quantitative evidence to separate the works into two distinct periods of production based on the different characteristics of their brushwork, demonstrating the usefulness of computer-based analysis as an added tool to help shed light on some standing debates among scholars.
Chapter 4

Distinguishing Norval Morrisseau’s Paintings via Curve Elegance

4.1 Introduction

The art forgery industry has become increasingly sophisticated to target the growing number of art collectors. Factories with assembly lines have been established to forge paintings from well-known artists [85]. Relatively skillful painters and modern technologies are involved in making counterfeits. Despite the use of modern technologies, such as carbon dating, lead dating, X-rays, multispectral imaging, and cross-section microscopy, authenticating visual art is still an open problem. A connoisseur can tell the authenticity of a painting by analyzing the emotions expressed by the artist. Authentic paintings often stimulate higher emotional responses than forgeries. Traditional painting authentication is a highly subjective and sophisticated appreciation process. Art historians utilize various heuristics and theories [39]. For instance, color, brushwork and composition are some important factors considered in artist attribution, dating, and painting style identification.

In computerized painting analysis, many problems lead to one main issue, that is, numerical characterization of paintings. The numerical features of paintings provide evidence for attribution and can be used for other purposes, e.g., retrieval. Existing work in the literature is mostly based on analyzing the characteristics

---

The materials in this chapter have appeared in [99].
of brushstrokes. Various traits of brushstrokes are by far the main subject for computational comparison of digitized fine art paintings, in large part due to the fact that brushstrokes have been considered to form unique “signatures” by art historians [39, 74, 86, 79]. The techniques of depiction, such as shading and glazing, suggest that texturelike features may be appropriate for brushstroke analysis [48, 39]. Although focusing on strokes diminishes the complexity of painting analysis, stroke extraction itself can be a daunting task, the difficulty of which varies tremendously among painters. How strokes were laid down in some paintings can be hard to discern even for human eyes.

However, some paintings, such as those by Norval Morrisseau (1932-2007), do not have clearly visible brushstrokes. In the case of Morrisseau, because he used primitive approaches to painting, specifically, modeling by light and shade does not exist, and edges are always crispy, it is relatively easy to extract the painted areas (which can be the result of one or multiple strokes). Through edge detection, the contour lines of the brushstrokes can be reliably extracted. In this study, contour curves acquired by edge detection are served as the main visual clue in order to capture the painters artistic characteristics. Measures of steadiness and neighborhood coherence have been developed and tested on the works of Norval Morrisseau, an aboriginal Canadian artist, as well as some known imitations. Figure 4.1 shows some examples in the dataset. Morrisseau, known as “Picasso of the North,” was arguably the greatest aboriginal artist ever to have lived in North America. His subject matter addressed the protection of the environment long before global warming entered our mainstream consciousness. The dataset includes 35 paintings, among which 19 are authentic paintings and 16 are forgeries confirmed by artist himself, which were photographed using both a digital SLR camera and a medium-format slide film camera.

4.2 Characterizing Elegance of Curves

A close observation of some works of Norval Morrisseau and several forgeries gives the impression that Morrisseau’s work is extraordinary in design and line quality. His lines appear both swift and steady, and the design of figures in his paintings appears harmonious and peaceful. While curves in the forgeries tend to be hesi-
Figure 4.1: Example authentic and counterfeit paintings from the dataset related to Norval Morrisseau, a native aboriginal painter from Canada.

tating and jagged, and the designs lack aesthetics. Hence, contour lines are found by edge detection to quantify such difference.

The EDISON edge detection algorithm [55] is employed to identify edge pixels for paintings. Input parameters of the algorithm are adjusted so that most of the contour lines are extracted by the program, while edges of low contrast such as within-stroke edges and noisy edges are not detected. An edge linking procedure is applied on edge maps to remove short edges and record coordinates of the pixels along curves in the right order (Figure 4.2).

Figure 4.2: Curve detection and linking applied to a painting.
4.2.1 Measure of Curve Steadiness

Figure 4.3 shows the edge map of a forgery. Many contour lines are jagged. The extent of jaggedness in the contour curves reflects the steadiness of the painter's hand. A painter of great draftsmanship, especially one that masters line work like Morriseau, distinguishes himself from unskillful painters in this quality.

The steadiness of curves is characterized using the extent of jaggedness. Figure 4.4 illustrates the basic ideas. The pixels of interest are highlighted by red circles. In the first case, the spanning angles of a corner point tend to decrease as the spanning edge increases. For a point at the tip of a spiky curve, the spanning angles tend to increase. While in the last case, for a point in a zigzag curve, the spanning angles tend to vary unmonotonically.

Denote a straight line segment connecting two points \( i \) and \( j \) by \( L_{i,j} \). At a given spanning length \( k \), to compute the spanning angle at edge point \( i \), \( L_{ik,i} \) and \( L_{i,i+k} \) are formed. The angle between the two is then called the spanning angle of
i at spanning length k. Let \( \beta_i^{(k)} \) represent the spanning angle for point i on a curve when the spanning length is k. The value of k is set to 3, 5, and 7 respectively. So there are three sequences of spanning angles for each curve. The measure for steadiness is defined as the ratio of points whose spanning angles across spanning lengths (or scales) do not vary monotonically. Again, let i be the index for points on a curve, and j the index for curves in a painting. Define

\[
\begin{align*}
  d'_i &= \beta_i^{(5)} - \beta_i^{(3)}, \\
  d''_i &= \beta_i^{(7)} - \beta_i^{(5)}.
\end{align*}
\]

(4.1)

Let \( s_i \) be the indicator whether \( d'_i \) and \( d''_i \) have the same sign. Then

\[
  r_j = \frac{\sum_i s_i}{\text{length}(j)},
\]

(4.2)

where \( \text{length}(j) \) is the length of the jth curve (number of points contained). Finally, the measure for the overall painting is the mean of \( r_j \):

\[
  m_1 = \text{mean}(r_j).
\]

(4.3)

### 4.2.2 Measures of Coherence by Tangent Directions

Two measures are defined to characterize the coherence of tangent directions of curves. The tangents along the detected curves are estimated approximately by the sum of the backward and forward vectors. If the spanning length is k, the backward vector is pointing from the \( (ik) \)th point to the ith point and the forward vector is from the ith point to the \( (i+k) \)th point. The tangent direction of the ith point is estimated by the sum of these two vectors. Similar to the color histogram, tangent angle histogram is used to describe tangent distribution of the curves in a painting. If the curves in a painting appear to flow in similar directions, the histogram tends to have high peaks in some bins. Otherwise, a random distribution of tangents is more likely to give a balanced histogram. In the tangent angle histogram, y axis represents the ratio of points whose tangent angles fall in the range of the bin. Since the exact tangent direction of majority points is unimportant, only the variance of the ratios are used to describe the distribution of tangents.

The aforementioned measure describes the flow of curves in a global way. The
third measure attempts to characterize tangent coherence locally. For each edge point, check other edge points in its neighborhood. If the tangent angle difference between the current point and the neighboring point is below a threshold, the neighboring point is claimed coherent with the current point. In this way, we can calculate the ratio of coherent neighboring points for any point on a curve, and use the average ratio as a coherence measure for the painting.

The first measure of coherence $m_2$ is to characterize the distribution of tangent angles. The angles range from $-180^\circ$ to $180^\circ$. The tangent angle histogram is divided evenly into 24 bins. The descriptor is the variance of the 24 bins in the histogram. For the second coherence measure, let $s_j$ represent the ratio of neighboring points which are coherent with the $j$th point, $t_j$ be the tangent angle, $n_j$ be the number of neighboring edge points, $T$ be the threshold of coherence which is set to be 15 in our experiment, and $n$ be the total number of points. Define

$$s_j = \frac{\sum_{k=1}^{n_j} I(|t_j - t_k| < T)}{n_j}, \quad m_3 = \frac{\sum_{j=1}^{n} s_j}{n}.$$  \hspace{1cm} (4.4)

The percentiles of coherence ratios (coherence measure II) are also used to demonstrate that authentic paintings have better neighborhood coherence. The $p\%$ percentile is a value that is greater than $p\%$ of the coherence ratios. Experimental results show that authentic works do have greater percentiles than forgeries, which means the coherence ratio distribution of authentic ones tends to be more left-skewed.

### 4.2.3 Classification Results

The steadiness and the coherence of contours are calculated for 35 digitized paintings in the dataset. Figure 4.5 shows the values of these measures. Clearly, most paintings of Morrisseau have better values for the three measures (i.e., smaller $m_1$ and greater $m_2$ and $m_3$).

Percentiles are calculated for the coherence ratios that are used to obtain the second coherence measure. In order to see the difference, a 95% 2-sample t-test is conducted for the 9 percentiles as well as the three measures mentioned above. The results for the t-tests are given in Table 4.1. The first column of the table shows
Figure 4.5: Calculated measures for the 19 Morrisseau paintings and 16 forgeries. Values in (a)-(c) are sorted.

different measures, among which 10% – 90% are the percentiles and \( m_1, m_2, m_3 \) are measures for steadiness and coherence. Figure 4.5(d) shows the 9 percentiles.

In order to further demonstrate the distinguishing power of the features, the 35 paintings are divided into four groups through random permutation of painting indices for cross-validation. Each of the first three groups has four forgeries and five authentic paintings while the fourth group has four forgeries and four authentic works. Features are normalized to have zero means and unit variances. SVM-light [3] is used for classification. Default parameters provided by the program are used for training. Each experiment involves one group as the test set and the other three as the training set. Table 4.2 provides the results of the cross-validation experiment.
Table 4.1: Results of 2-sample t-test

<table>
<thead>
<tr>
<th>Forgery</th>
<th>Genuine</th>
<th>Difference</th>
<th>P – value</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>0.10326</td>
<td>0.13096</td>
<td>-0.027697</td>
</tr>
<tr>
<td>20%</td>
<td>0.17271</td>
<td>0.22356</td>
<td>-0.050856</td>
</tr>
<tr>
<td>30%</td>
<td>0.23864</td>
<td>0.3052</td>
<td>-0.066522</td>
</tr>
<tr>
<td>40%</td>
<td>0.3061</td>
<td>0.3843</td>
<td>-0.078155</td>
</tr>
<tr>
<td>50%</td>
<td>0.3790</td>
<td>0.4613</td>
<td>-0.082282</td>
</tr>
<tr>
<td>60%</td>
<td>0.4574</td>
<td>0.5396</td>
<td>-0.082218</td>
</tr>
<tr>
<td>70%</td>
<td>0.5426</td>
<td>0.6290</td>
<td>-0.086442</td>
</tr>
<tr>
<td>80%</td>
<td>0.6467</td>
<td>0.7280</td>
<td>-0.081236</td>
</tr>
<tr>
<td>90%</td>
<td>0.7895</td>
<td>0.8628</td>
<td>-0.073310</td>
</tr>
</tbody>
</table>

\( m_1 \) 0.3403 0.3177 0.022556 0.000
\( m_2 \) 0.000299 0.000589 -0.000289 0.000
\( m_3 \) 0.41514 0.47846 -0.063318 0.000

Table 4.2: SVM cross-validation error rates

<table>
<thead>
<tr>
<th>Test group</th>
<th>Error on training set</th>
<th>Error on test set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>misclassified/total</td>
<td>misclassified/total</td>
</tr>
<tr>
<td>1</td>
<td>1/26 or 3.8%</td>
<td>2/9 or 22%</td>
</tr>
<tr>
<td>2</td>
<td>4/26 or 15%</td>
<td>0/9 or 0%</td>
</tr>
<tr>
<td>3</td>
<td>3/26 or 12%</td>
<td>0/9 or 0%</td>
</tr>
<tr>
<td>4</td>
<td>2/27 or 7.4%</td>
<td>2/8 or 25%</td>
</tr>
</tbody>
</table>

4.3 Summary

This chapter presents an approach to characterize steadiness and coherence of contour lines in paintings. Existing techniques on painting authentication have been primarily based on brushstroke characteristics and cannot be applied to certain styles of paintings when individual brushstrokes are not clearly visible. The steadiness of the contour lines reflects the draftsmanship of a painter in line work. And the proposed coherence measures can be used to distinguish the authentic works from the forgeries. Paintings by Morisseau and forgeries are analyzed. It is found that Morisseau’s paintings consistently demonstrate higher level of steadiness and coherence in curves. Whereas the techniques demonstrate their power on this dataset, they may be insufficient in identifying skillfully forged paintings. A study involving forgeries from additional sources will be desired.
5.1 Introduction

One of the most important factors that distinguish professional photographs and snapshots is the composition each one has. Although the apparatuses make difference, a photographer who masters the art of composition can still produce more appealing pictures with an inexpensive setup than one with high-end equipment but knowing little about composition. The challenge lies in the way of “seeing” through the viewfinder. Therefore computational studies on automatic analysis of photography composition can potentially benefit amateur photographers by offering improvement on the “seeing” skills.

Composition is the art of putting things together with conscious thoughts. In photography, it concerns the arrangement of various visual elements, such as line, color, space, etc. Composition is closely related to the aesthetic qualities of photographs. In the past few years, studying photography composition especially under the consideration of aesthetics has drawn particular interests from researchers in image processing and computer vision. The potential applications for studies in this category range from aesthetics assessment to aesthetics enhancement. Researchers have adopted some classic composition principles in their studies. The materials in this chapter have appeared [100].
most popular ones include “rule of thirds”, “simplicity”, “visual balance”, “color harmony”, etc.

In this work, photography composition is studied from the perspective of spatial layout, which is about how visual elements are geometrically arranged in a picture. General photographs are classified into five pre-selected spatial composition categories. Thus far, very little research has been done to address the problem of understanding the geometric organization of visual elements in photographs. The composition classifier is further integrated into the SIMPLIcity image retrieval system [90] so that retrieved images will tend to have similar compositions. A conventional image retrieval system returns images according to visual similarity. However a photographer’s intention is more likely to search for pictures by composition rather than by visual details. Furthermore, the new system also provides the option to rank retrieved images by their aesthetic ratings. Therefore, the system is able to provide on-site feedback via high-quality photographs with similar compositions and offer users the opportunity of “learning photography through exemplars” on the spot.

It is important to point out that the principles and rules in photography should be intended more as guideline for beginners than universal laws. Ansel Adams once said: “There are no rules for good photographs, there are only good photographs.” Indeed, many great works have been created that do not follow one rule and another. The rules are established on the rationale behind them. For example, “rule of thirds” is merely one simple way to make good use of space. To professional artists, the concerns are the underlying rationales, not the rules. Therefore tools that are developed based on the pre-determined rules are mainly targeted at amateurs, but not for professionals.

5.2 Spatial Composition Categorization

These typical spatial layouts, namely “horizontal”, “vertical”, “centered”, “diagonal”, and “textured”, are identified based on guiding principles in photography. According to long-existing art and photography principles, lines formed by linear elements are important because they lead the eye through the image and contribute to the mood of the photograph. Horizontal, vertical, and diagonal lines are
associated with serenity, strength, and dynamism respectively [25, 26]. Therefore “horizontal”, “vertical”, and “diagonal” are included as three composition categories. Photographs with a centered main subject and a clear background fall into the category “centered”. “Textured” photographs refer to those image appear like a patch of texture or a relatively homogeneous pattern, for example, an image of a brick wall. Figure 5.1 displays typical photo examples for the composition types under consideration.

The five categories of composition are not mutually exclusive. For example, a photograph may contain a diagonal element and meanwhile has a dominant horizontal line across the frame. Therefore the categorization problem can also be considered as a detection problem, i.e. to identify dominant compositional elements if there are some in the photographs. Several classifiers are applied sequentially to an image: “textured” versus “non-textured”, “diagonal” versus “non-diagonal”, and finally a possibly overlapping classification of “horizontal”, “vertical”, and “centered”. For example, an image can be classified as “non-textured”, “diagonal”, and “horizontal”. A method in [90] is employed to classify “textured” and “non-textured” images. The method is based on the intuition that colors in textured images usually scatter over the entire image, while non-textured images often have
clumped regions. It has been demonstrated that retrieval performance can be improved for both textured and non-textured images by first classifying them [90]. The other two classifiers are developed in this work, with details to be presented shortly.

5.2.1 Diagonal Element Detection

Diagonal elements are strong compositional constituents. The diagonal rule in photography states that a picture will appear more dynamic if the objects fall or follow a diagonal line. Photographers often use diagonal elements as the visual path to draw viewers’ eyes through the image [29]. Visual path means the path of eye movement while a viewer is looking at a photograph [92]. When such a visual path stands out in the picture, it also has the effect of uniting individual parts in a picture. The power of the diagonal lines for composition was exploited very early on by artists. Speed [77] discussed in great details how Velazquez used the diagonal lines to unite a picture in his painting “The Surrender of Breda”.

Because of the importance of diagonal visual paths for composition, a spatial composition category is created for diagonally composed pictures. More specifically, there are two subcategories, diagonal from upper left to bottom right and from upper right to bottom left. A photo is considered to have a diagonal composition if diagonal visual paths exist.

Detecting the exact diagonal visual paths is difficult. Segmented regions or edges provided by the usual image processing techniques often can only serve as ingredients, aka, local patterns, either because of the nature of the picture or the limitation of the processing algorithms. In contrast, an element refers to a global pattern, e.g., a fragmented straight line (corresponding to multiple edge segments) that has presence in a large area of the image plane.

The algorithm for detecting diagonal visual paths is designed according to the following principles. These principles present here are for the diagonal case, but they apply similarly to other directional visual paths.

1. Principle of multiple visual types: Lines are effective design elements in creating compositions, but “true” lines rarely exist in real world. Lines we perceive in photographs usually belong to one of these types: outlines of
forms; narrow forms; lines of arrangement; and lines of motion or force [24]. Diagonal elements are not restricted to actual diagonal lines of an image plane. They could be the boundary of a region, a linear object, and even an imaginary line along which different objects align. Linear objects, such as pathways, waterways, and the contour of a building, can all create visual paths in photographs. When placed diagonally, they are generally perceived as more dynamic and interesting than other compositions. Figure 5.2 shows examples of using diagonal compositions in photography.

2. **Principle of wholes/Gestalt Law**: Gestalt psychologists studied early on the phenomenon of human eyes perceiving visual components as organized patterns or wholes, known as the Gestalt law of organization. According to the Gestalt Law, the factors that aid in human visual perception of forms include Proximity, Similarity, Continuity, Closure and Symmetry [78].

3. **Principle of tolerance**: Putting details along diagonals creates more interesting compositions. Visual elements such as lines and regions slightly off the ideal diagonal direction can still be perceived as diagonal and are usually more natural and interesting [94].

4. **Principle of prominence**: A photograph can contain many lines, but dominant lines are the most important in regard to the effect of the picture [30, 27]. Visual elements need sufficient span along the diagonal direction in order to strike a clear impression.

![Figure 5.2: Photographs of diagonal composition.](image)

Following the above principles, diagonal ingredients are first found from low-level visual cues using both regions obtained by segmentation and connected lines obtained by edge detection. Then, according to the Gestalt Law, the ingredients
are merged into elements, i.e., more global patterns. The prominence of each merged entity is then assessed. Next, the algorithms for detecting diagonal visual paths using segmented regions and edges will be described respectively.

5.2.1.1 Diagonal Segment Detection

Image segmentation is often used to simplify the image representation. It can generate semantically meaningful regions that are easier for analysis. This section describes the approach to detecting diagonal visual paths based on segmented regions. The recently developed image segmentation algorithm [47] is used for it achieves state-of-the-art accuracy at a speed sufficiently fast for real-time systems. The algorithm also ensures that the segmented regions are spatially connected, a desirable trait many algorithms do not possess.

After image segmentation, the orientation of the moment axis of each segment is calculated and taken as the orientation of the segment. The moment axis is the direction along which the spatial locations of the pixels in the segment have maximum variation. It is the first principal component direction for the data set containing the coordinates of the pixels. For instance, if the segment is an ellipse (possibly tilted), the moment axis is simply the long axis of the ellipse. The orientation of the moment axis of a segmented region measured in degrees is computed according to [73]. Suppose the number of pixels in segment $S$ is $s$. The coordinates of pixels are denoted by $(x, y)$. For each segment, the following three quantities are computed,

$$m_x = \sum_{(x,y) \in S} x^2 - \frac{(\sum_{(x,y) \in S} x)^2}{s},$$

$$m_y = \sum_{(x,y) \in S} y^2 - \frac{(\sum_{(x,y) \in S} y)^2}{s},$$

$$m_{xy} = \sum_{(x,y) \in S} xy - \frac{\sum_{(x,y) \in S} x \sum_{(x,y) \in S} y}{s}.$$
Then the orientation $e$ of segment $S$ is defined as

$$e = \arctan \frac{m_x - m_y + \sqrt{(m_x - m_y)^2 + 4 \times m_{xy}^2}}{2 \times m_{xy}}. \quad (5.1)$$

Next, certain segmented regions are merged according to the Gestalt Law to form visual elements. Currently, only a simple case of disconnected visual path is considered, where the orientations of all the disconnected segments are diagonal.

Here first introduces a few notations before describing the rules for merging.

\[
\begin{align*}
\vec{v}_d &: \text{the normalized column vector of the diagonal direction.} \\
\vec{v}_d^\perp &: \text{the orthogonal direction of } \vec{v}_d. \\
S &: \text{a segmented region.} \\
\vec{x} &: \text{coordinates of pixels in segment } S, \text{i.e. } \vec{x} = (x_h, x_v)^t. \\
\vec{x} \cdot \vec{v} &: \text{the projection of a pixel } \vec{x} \text{ onto the direction whose normalized vector is } \vec{v}. \\
P(S, \vec{v}) &: \text{a set containing the projected coordinates of all the pixels in } S, \text{i.e. } P(S, \vec{v}) = \{ \vec{x} \cdot \vec{v} : \vec{x} \in S \}. \\
|P(S, \vec{v})| &: \text{the length (also called spread) of the projection, that is, the range of values in the projected set, i.e.} |P(S, \vec{v})| = \max_{\vec{x}_i, \vec{x}_j \in S} |\vec{x}_i \cdot \vec{v} - \vec{x}_j \cdot \vec{v}|.
\end{align*}
\]

The rules for merging, called “Similarity”, “Proximity”, and “Continuity”, are listed below. Two segments satisfying all of the rules are merged.

**Similarity**: Two segments $S_i$, $i = 1, 2$, with orientations $e_i$, $i = 1, 2$, are similar if the following criteria are satisfied:

1. Let $[\hat{\varphi}, \hat{\varphi}]$ be the range for nearly diagonal orientations. $\hat{\varphi} \leq e_i \leq \hat{\varphi}$, $i = 1, 2$. That is, both $S_1$ and $S_2$ are nearly diagonal.

2. The orientations of $S_i$, $i = 1, 2$, are close:

$$|e_1 - e_2| \leq \beta,$$

where $\beta$ is a pre-chosen threshold.
3. The lengths of $P(S_i, \vec{v}_d)$, $i = 1, 2$, are close:

$$r = \frac{|P(S_1, \vec{v}_d)|}{|P(S_2, \vec{v}_d)|}, \quad r_1 \leq r \leq r_2,$$

where $r_1 < 1$ and $r_2 > 1$ are pre-chosen thresholds.

- **Proximity**: Segments $S_i$, $i = 1, 2$, are proximate if their projections on the diagonal direction, $P(S_i, \vec{v}_d)$, $i = 1, 2$, are separated by less than $p$, and the overlap of their projections is less than $q$.

- **Continuity**: Segments $S_i$, $i = 1, 2$, are continuous if their projections on the direction orthogonal to the diagonal, $P(S_i, \vec{v}_c)$, $i = 1, 2$, are overlapped.

The various thresholds in the three rules are chosen as below.

1. $\beta = 10^\circ$.

2. $r_1 = 0.8$, $r_2 = 1.25$.

3. The values of $p$ and $q$ are decided adaptively according to the sizes of $S_i$, $i = 1, 2$. Let the spread of $S_i$ along the diagonal line be $\lambda_i = |P(S_i, \vec{v}_d)|$. Then $p = k_p \min(\lambda_1, \lambda_2)$ and $q = k_q \min(\lambda_1, \lambda_2)$, where $k_p = 0.5$ and $k_q = 0.8$.

The value of $p$ determines the maximum gap allowed between two disconnected segments to continue a visual path. The wider the segments spread over the diagonal line, the more continuity they present to the viewer. Therefore, heuristically, a larger gap is allowed, which is why $p$ increases with the spreads of the segments. On the other hand, $q$ determines the extent of overlap allowed for the two projections. By a similar rationale, $q$ also increases with the spreads. If the projections of the two segments overlap too much, the segments are not merged because the combined spread of the two differs little from the individual spreads.

4. The angular range $[\hat{\varphi}, \hat{\varphi}]$ for nearly diagonal orientations is determined adaptively according to the geometry of the rectangle bounding the image.

After studying many works of visual art, the Dutch photographer Edwin Westhoff discovered that artists often placed important details on the diagonals of a
square to draw viewers’ attention [94]. A practical suggestion on photography composition is made in an online article, where boundary lines are drawn using the sixth points on the borders of the image plane. A sixth point along the horizontal border from the upper left corner locates on the upper border and is away from the corner by one-sixth of the image width. Other sixth (or third) points from any corner and either horizontally or vertically can be defined similarly.

Suppose we look for an approximate range for the diagonal direction going from the upper left corner to the bottom right. The sixth and third points with respect to the two corners are found. As shown in Figure 5.3 (a), these special points are used to create two stripes marked by lime and blue colors respectively. Let the orientations of the lime stripe and the blue stripe in Figure 5.3 (a) be \( \varphi_1 \) and \( \varphi_2 \). Set \( \check{\varphi} = \min(\varphi_1, \varphi_2) \), and \( \hat{\varphi} = \max(\varphi_1, \varphi_2) \). A direction \( \vec{v} \in [\check{\varphi}, \hat{\varphi}] \) is claimed nearly diagonal. The angular range for the diagonal direction from the upper right corner to the bottom left can be obtained in a similar way. The reason to use the orientations of the stripes instead of nearly diagonal bounding lines is that when the width and the height of an image are not equal, the orientation of a stripe will twist toward the elongated side to some extent.

![Figure 5.3: Diagonal orientation bounding conditions.](image)

From now on, a segment can be a merged entity of several segments originally given by the segmentation algorithm. For brevity, the merged entity is still called a segment. Applying the principle of tolerance, a segment is filtered out from diagonal if its orientation is outside the range \( [\check{\varphi}, \hat{\varphi}] \), the same rule applied to the smaller segments before merging.

After removing non-diagonal segments, the remaining segments are examined for their prominence. At last, only segments with a significant spread along the diagonal direction are retained according to the principle of prominence. For seg-

---

\(^2\)http://www.colorpilot.com/comp_rules.html
ment $S$, if $|\mathcal{P}(S, \vec{v})| \geq k_1 \times l$, where $l$ is the length of the diagonal line and $k_1 = \frac{2}{3}$ is a threshold, the segment is declared a diagonal visual path. It is observed that a diagonal visual path is often a merged entity of several small segments originally produced by the segmentation algorithm, which are not prominent individually.

5.2.1.2 Diagonal Edge Detection

According to the principle of multiple visual types, besides segmented regions, lines and edges can also form visual paths. Moreover, segmentation can be unreliable sometimes because over-segmentation and under-segmentation often cause diagonal elements to be missed. Among photographs showing diagonal composition, many contain linear diagonal elements. Those linear diagonal elements usually have salient boundary lines along the diagonal direction, which can be found through edge detection. Therefore edges are used as another visual cue, and the results obtained based on both edges and segments are combined to increase the sensitivity of detecting diagonal visual path.

The Edison edge detection algorithm [55] is used to detect edge pixels. It has been experimentally demonstrated that Edison edge detection can generate cleaner edge maps than many other methods [55]. All the edges are examined and only those oriented diagonally and significant enough are considered as a visual path.

Based on the same set of principles, the whole process of finding diagonal visual paths based on edges is similar to the detection of diagonal segments. The major steps are described below. An edge is denoted by $E$, which is a set of coordinates of pixels located on the edge. As with segments, the notation $\mathcal{P}(E, \vec{v})$ is used for the projection of $E$ on a direction $\vec{v}$.

1. **Remove non-diagonal edges**: First, edges outside the diagonal stripe area, as shown in Figure 5.3 (b), are excluded. Secondly, for every edge $E$, its orientation is computed using equation (5.1). Edges with orientation $e$ such that $e \notin [\zeta_1, \zeta_2]$ are regarded as non-diagonal edges and filtered out. The choice of $\zeta_1$ and $\zeta_2$ will be discussed later.

2. **Merge edges**: After removing non-diagonal edges, short edges along the diagonal direction are merged into longer edges. The merging criterion is similar to the proximity and continuity rules used for diagonal segments.
Two edges are merged if: 1) their projections onto the diagonal line are close to each other but not excessively overlapped; and 2) their projections onto the orthogonal direction \( \vec{c} \) are also close to each other.

3. **Examine prominence**: Edges formed after the merging step are examined for their spread along the diagonal direction. For every edge \( E \), compute the spread of the projection \( |P(E, \vec{v}_d)| \). An edge \( E \) is taken as a diagonal visual element if \( |P(E, \vec{v}_d)| \geq \xi \), where \( \xi \) is a threshold to be described next.

The values of thresholds \( \zeta_1, \zeta_2 \) and \( \xi \) are determined by the size of a given image. \( \zeta_1 \) and \( \zeta_2 \) are used to filter out edges whose orientations are not quite diagonal, and \( \xi \) is used to select edges that spread widely along the diagonal line. The third points on the borders of the image plane are used to set bounding conditions. Figure 5.3 (c) shows two lines marking the angular range allowed for a nearly diagonal direction from the upper left corner to the lower right corner. Both lines in the figure are off the ideal diagonal direction to some extent. Let \( \zeta_1 \) and \( \zeta_2 \) be their orientations, and \( \xi_1 \) and \( \xi_2 \) be their spreads over the diagonal line. The width and height of the image are denoted by \( w \) and \( h \). By basic geometry, \( \xi_i, i = 1, 2 \) can be calculated using the formulas:

\[
\xi_1 = \frac{h^2 + 3w^2}{3\sqrt{h^2 + w^2}}, \quad \xi_2 = \frac{3h^2 + w^2}{3\sqrt{h^2 + w^2}}.
\]

The threshold \( \xi \) is then set by \( \xi = \min(\xi_1, \xi_2) \).

5.2.2 “Horizontal”, “Vertical” and “Centered” Composition Types

This section presents the method for differentiating the last three composition categories: “horizontal”, “vertical” and “centered”. Photographs belonging to each of these categories have distinctive spatial layouts. For instance, a landscape with blue sky at the top and field at the bottom conveys a strong impression of horizontal layout. Images from a particular category usually have some segments that are characteristic of that category, e.g., a segment lying laterally right to left for “horizontal” photographs, and a homogeneous background for “centered” photographs.
In order to quantitatively characterize spatial layout, the spatial relational vector (SRV) of a region is defined to specify the geometric relationship between the region and the rest of the image. The spatial layout of the entire image is then represented by the set of SRVs of all the segmented regions. The dissimilarity between spatial layouts of images is computed by the IRM distance [49]. Ideally the spatial layout of photographs means the relationship between each semantically meaningful object and its surrounding space. However, object extraction is inefficient and extremely difficult for photographs in general domain, regions obtained by image segmentation algorithms are used instead as a reasonable approximation. Moreover, for painters, reducing the complicated appearance into simple masses is a necessary step in her composition, and expresses the “essense” of a painting’s structure [77].

5.2.2.1 Spatial Relational Vectors (SRV)

The SRV is proposed to characterize the geometric position and the peripheral information about a pixel or a region in the image plane. It is defined at both pixel-level and region-level. When computing the pixel-level SRV, the pixel is regarded as the reference point, and all the other pixels are divided into 8 zones by their relative positions to the reference point. If the region that contains the pixel is taken into consideration, SRV is further differentiated into two modified versions, inner SRV and outer SRV. The region-level inner (outer) SRV is obtained by averaging pixel-level inner (outer) SRVs over the region. Details about SRV implementation will be given next. As we will see, SRV is scale-invariant, and depends on the spatial position and the shape of the segment.

At a pixel with coordinates \((x, y)\), 4 lines passing through it are drawn. As shown in Figure 5.4 (a), the angles between adjacent lines are equal and stride symmetrically over the vertical, horizontal, 45° and 135° lines. Denote the 8 angular areas of the plane by “UPPER”, “UPPER-LEFT”, “LEFT”, “BOTTOM-LEFT”, “BOTTOM”, “BOTTOM-RIGHT”, “RIGHT”, and “UPPER-RIGHT” zones respectively. The SRV of pixel \((x, y)\) summarizes the angular positions of all the other pixels with respect to \((x, y)\). Specifically, the area percentage \(v_i\) of each zone, \(i = 0, ..., 7\), is calculated with respect to the whole image and the pixel-level SRV \(V_{x,y}\) is then formulated by \(V_{x,y} = (v_0, v_1, ..., v_7)^t\).
The region-level SRV is defined in two versions called respectively inner SRV, denoted by \( V' \), and outer SRV, denoted by \( V'' \). The image plane can be divided into 8 zones at any pixel in a region by the above scheme. As shown in Figure 5.4 (b), for each of the 8 zones, some pixels are inside the region and some are outside. Depending on whether a pixel belongs to the region, the 8 zones are further divided into 16 zones. Those zones within the region are called inner pieces and those outside outer pieces. Area percentages of the inner (or outer) pieces with respect to the area inside (or outside) the region form the inner SRV \( V'_{x,y} \) (or outer SRV \( V''_{x,y} \)) for pixel \((x,y)\).

The region-level SRV is defined as the average of pixel-level SRVs for pixels in that region. The outer SRV \( V''_{R} \) of a region \( R \) is \( V''_{R} = \frac{\sum_{(x,y) \in R} V''_{x,y}}{m} \), where \( m \) is the number of pixels in region \( R \). In practice, to speed up the calculation, the pixels \((x,y)\) in \( R \) can be subsampled and \( V''_{R} \) is computed by averaging over only the sampled pixels. If a region is too small to occupy at least one sampled pixel according to a fixed sampling rate, the pixel at the center of the region is used to compute \( V''_{R} \), i.e. \( V''_{R} \) is set to be the outer SRV for the center of the region.

The outer SRV is used to characterize the spatial relationship of a region with respect to the rest of the image. Then an image with \( N \) segments \( R_i, i = 1, ..., N \), can be described by \( N \) region-level outer SRVs, \( V''_{R_i}, i = 1, ..., N \), together with the area percentages of \( R_i \), denoted by \( w_i \). In summary, an image-level SRV descriptor is a set of weighted SRVs: \( \{(V''_{R_i}, w_i), i = 1, ..., N\} \). This descriptor is called the spatial layout signature.

![Figure 5.4: Division of the image into 8 angular areas with respect to a reference pixel.](image)
5.2.2.2 “Horizontal”, “Vertical” and “Centered” Composition Classification

The k-nearest neighbor algorithm (k-NN) is used to classify the three composition categories: “horizontal”, “vertical” and “centered”. Inputs to the k-NN algorithm are the spatial layout signatures of images. The training dataset includes equal number of manually labeled examples in each category. In the experiment, the sample size for each category is 30. The distance between the spatial layout signatures of two images is computed using the IRM distance [49]. IRM algorithm is a greedy implementation of the Mallows distance. It allocates a weight to any pair of regions, one from each image. The weight measures the significance of matching the two regions. A greater value of the weight means the distance between the two regions contributes more to the image-level distance than a region pair with smaller matching weight. Suppose two images in comparison have $N_1$ and $N_2$ regions respectively. Let $V_{R_i}^{n(1)}$, $i = 1, ..., N_1$ and $V_{R_j}^{n(2)}$, $j = 1, ..., N_2$ denote the outer SRVs of regions in the two images. $d_{i,j}$ is the Euclidean distance between $V_{R_i}^{n(1)}$ and $V_{R_j}^{n(2)}$, and $s_{i,j}$ is the weight allocated by IRM for the two regions. Then the image-level distance $D$ is calculated by:

$$D = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} d_{i,j} \times s_{i,j}.$$ 

Given an image, its $k$ nearest neighbors in the training dataset can be found using this distance measure. The image is assigned to the category which has the most votes.

5.3 Composition Sensitive Image Retrieval

The classic approach taken by many image retrieval systems is to measure the visual similarity based on low-level features. A large family of visual descriptors have been proposed in the past to characterize images from the perspectives of color, texture, shape, etc. However, due to the fact that many visual descriptors are generated by local feature extraction processes, the overall spatial composition of the image is usually lost. In semantic content oriented applications, spatial lay-
out information of an image may not be critical, but for photography, the overall spatial composition can be a major factor affecting how an image is perceived. For photographers, it can be more interesting to search for photos with similar composition style rather than visual details. The algorithms described previously can capture strong compositional elements in photos and classify them into six composition categories, with five main categories namely “textured”, “horizontal”, “vertical”, “centered”, and “diagonal”, and the “diagonal” category is further subdivided into two categories “diagonal_ulbr” (upper left to bottom right) and “diagonal_urbl” (upper right to bottom left). The composition classification is used in the retrieval system to return images with similar composition.

The SIMPLlicity system [90] is used to retrieve images with similar visual content. The top $K$ retrieved images is then re-ranked by their spatial composition and aesthetic scores. SIMPLlicity is a semantic-sensitive region based image retrieval system. It partitions an image into blocks with $4 \times 4$ pixels and extracts a six dimensional feature vector for each block. The blockwise feature includes average colors in LUV space and energy measures in high frequency bands of wavelet transform. It then uses a K-means based algorithm to cluster these blocks into homogeneous regions. The blockwise features are averaged out within a region to form the regionwise feature of color and texture. Normalized inertia of order 1 to 3 are used as region shape descriptors. Distance between regions is defined on texture and color features for textured images and adjusted by shape features for non-textured images. IRM is used to measure visual similarity between images. In the proposed system, the rank of an image is determined by three factors: visual similarity, spatial composition categorization, and aesthetic score. Since these factors are of different modality, a ranking schema is employed rather than a complicated scoring equation.

Given a query, SIMPLlicity first retrieves $K$ images and gives an initial ranking. When composition is taken into consideration, images with the same composition categorization as the query will be moved to the top of the ranking list.

The composition classification is non-exclusive in the context of image retrieval. For instance, a “textured” image can still be classified into “horizontal”, “vertical” or “centered”. The classification results obtained from the classifiers is coded in a six-dimensional vector $c$, corresponding to six categories (“diagonal” has two sub-
categories “diagonal_ulrb” and “diagonal_urbl”). Each dimension records whether the image belongs to a particular category, with 1 being yes and 0 no. Note that an image can belong to multiple classes generated by different classifiers. The image can also be assigned to one or more categories among “horizontal”, “vertical” and “centered” if neighbors belonging to the category found by k-NN reach a substantial number (currently $k/3$ is used). Non-exclusive classification is more robust than exclusive classification in practice because a photograph may be reasonably assigned to more than one compositional category. Non-exclusive classification can also reduce the negative effect of misclassification into one class. Figure 5.5 shows example pictures that are classified as more than one category.

![Example pictures](image1.png)

Figure 5.5: Photographs classified into multiple categories. “diagonal_ulbr” represents the diagonal from the upper left corner to the bottom right corner, and “diagonal_urbl” represents the other.

The compositional similarity between the query image and another image can be defined as

$$s_i = \sum_{k=0}^{3} I(c_{qk} = c_{ik} \text{ and } c_{qk} = 1) + 2 \times \sum_{k=4}^{5} I(c_{qk} = c_{ik} \text{ and } c_{qk} = 1),$$

where $c_q$ and $c_i$ are categorization vectors for the query image and the other image, and $I$ is the indicator function returning 1 when the input condition is true, 0 otherwise. The last two dimensions of the categorization vector correspond to the two diagonal categories. The matching function is multiplied by 2 to encourage matching of diagonal categories in practice. Note that the value of $s_i$ is between 0 and 7, because one image can at most be classified into 5 categories, which are “textured”, “diagonal_ulbr”, “diagonal_urbl” and two of the other three. Therefore
by adding composition classification results, the $K$ images are divided into 8 groups corresponding to compositional similarity from 0 to 7. The original ranking based on visual similarity remains within each group. When aesthetic rating is further introduced into the ranking schema, images within each group are reordered by aesthetic ratings. Let $r_i$, $s_i$ and $q_i$ denote the rank, compositional similarity, and aesthetic score of image $i$. The ranking schema can be expressed as:

$$r_i \prec r_j \text{ if } \begin{cases} s_i > s_j \\ q_i > q_j, \ s_i = s_j. \end{cases}$$

The reason for using such a ranking scheme is that the three perspectives incorporated are of different modalities and it is difficult to put these distinct measurements in the same space. Although the composition analysis is performed on the results returned by a CBIR system SIMPLIcity, the influence of this component in the retrieval process can be modified by adjusting the number of images $K$ returned by SIMPLIcity. This provides flexibility for the user to vary her focus on either composition or visual similarity. For example, a large $K$ will retrieve more compositionally relevant photographs, and meanwhile reduce the importance of content similarity. Experiment shows that in most cases the retrieved results become stable for our dataset when $K > 300$, a value expected to vary with the size of dataset. Figure 5.6 provides some examples showing how different values of $K$ can affect the retrieved results.

### 5.4 Evaluation of Spatial Layout Classifiers

The spatial layout categorization algorithms described in Section 5.2 analyze the compositional properties of a photograph. Three classifiers are applied on each image, namely the “textured” vs “non-textured” classifier, the diagonal element detector, and the k-NN classifier for “horizontal”, “vertical” and “centered” compositions. The classification can be either exclusive or non-exclusive. Exclusive classification will be used to evaluate the performance of classifiers, while non-exclusive classification is chosen for the retrieval system. This section discusses experimental results in diagonal element detection, and k-NN classification for “horizontal”, “vertical”, and “centered” classes.
Figure 5.6: Images retrieved by different values of $K$, using composition categorization to rerank results returned by SIMPLIcity. (a) The query image is at the top center and top 8 re-ranked images retrieved when $K = 50$ (first row), $K = 100$ (second row) and $K = 300$ (third row) are shown in three rows; (b) A second example with the query image at the top center and the top 8 re-ranked images retrieved when $K = 50$ (first row), $K = 100$ (second row) and $K = 200$ (third row) are shown in three rows.
5.4.1 Diagonal Element Detection

Algorithms for detecting diagonal element are provided in Section 5.2.1. Both segments and edges are used as visual cues. After segmentation and edge detection, small segments or edges aligning along the same diagonal direction are merged. These merged segments or edges with wide spread along either of the diagonal lines are marked as diagonal elements. The images which contain diagonal elements are classified into the “diagonal” category. Figure 5.7 shows some examples for merging segments or edges. Images in the second column show edges and segments generated by edge detection and image segmentation, where disconnected edges or segments are marked by different colors. The third column contains images displaying the detected diagonal elements after merging.

The current algorithm has some of the following limitations. Firstly, diagonally oriented edges in a noisy context (many other edges in the neighborhood) do not emerge as salient visual elements. Merging edges in a noisy context will lead to false identification of diagonal composition. Secondly, the algorithm only merges diagonally oriented edges or segments. Therefore, it cannot detect more subtle diagonal visual paths formed by edges or segments that are not individually oriented in a diagonal fashion.

Some examples of diagonal element detection are given in Figure 5.8. Images in the first column are the original pictures. The second column and the third column contain image segmentation and diagonal segment detection results, while the last two columns show the edge detection and diagonal edge detection results. Currently, the system provides only information about whether a photograph is diagonally composed or not. The system can be enhanced to take requests as to whether a photograph contains near-diagonal elements so that users can be alerted to adjust the frame to achieve a stronger diagonal composition while taking photographs.

5.4.2 Classification of Spatial Composition Categories

“Horizontal”, “Vertical” and “Centered”

The k-NN classification algorithm ($k = 30$) is applied on the spatial layout signatures of images using the IRM distance. In the training dataset, the sample size
Figure 5.7: Merging edges or segments to form more prominent diagonal elements. First column: original images; Second column: edge detection or segmentation results; Third column: edges or segments that are merged and identified as diagonal elements.

for each category is 30. Figure 5.9 shows example photos classified as “horizontal”, “vertical” and “centered” by the k-NN classifier. Misclassification can be caused by a biased training dataset because the image samples in the training dataset may not represent sufficiently the corresponding category. Poor segmentation can also lead to misclassification, since the spatial layout signature is sensitive to the image segmentation results. For example, “centered” photographs can be misclassified if the background is incorrectly broken into multiple segments. We notice that the spatial layout signature of a “centered” photo distinguishes itself clearly from “horizontal” and “vertical” signatures when the centered segment, usually the subject,
fills a major portion of the image plane. It can also occur when the background segment takes up a major portion of the image plane or when the segments on the boundary region are evenly distributed.

5.4.3 Confusion Table

In order to evaluate the performance of the composition classification algorithms, the three classifiers are applied on 222 manually labeled photographs, among which 50 are horizontally composed, 51 are vertically composed, 50 are centered, and 71 are diagonally composed (35 have visual elements along the diagonal line from the upper left corner to the bottom right corner, and 36 have diagonal elements in the other direction). Images in the testing dataset are pictures which have composition clarity and fit into single category.
This experiment performs an exclusive classification. The classifiers are applied on an image sequentially. That is, the “textured” vs “non-textured” classifier is first used to determine whether the image looks “textured”. If this classifier labels it “non-textured”, the diagonal element detector is then applied. Provided any diagonal element is detected, the image is assigned to one of the diagonal categories according to its orientation. If the category of the image is still undetermined, the k-NN classifier finally decides its composition type by classifying it under the
Table 5.1: Confusion table for composition classification on 222 images from five categories

<table>
<thead>
<tr>
<th></th>
<th>h</th>
<th>v</th>
<th>c</th>
<th>ulbr</th>
<th>urbl</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>44</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>v</td>
<td>0</td>
<td>34</td>
<td>7</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>c</td>
<td>6</td>
<td>3</td>
<td>29</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>ulbr</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td>urbl</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>33</td>
</tr>
</tbody>
</table>

category which gets the most votes. This experiment only evaluates the diagonal detector and k-NN classifier for “horizontal”, “vertical” and “centered”, since “textured” vs “non-textured” classifier is existing work from [90]. Hence the first step of the above process is skipped in this experiment. Denote these categories by “h” for “horizontal”, “v” for “vertical”, “c” for “centered”, “ulbr” for diagonal direction from the upper left corner to the bottom right corner, and “urbl” for the other. Table 5.1 gives the confusion table for this testing dataset.

5.4.4 User Perception on Composition Layout

In this section, the performance of the composition categorization algorithms is evaluated in a user study. Around 30 students were recruited to participate, most of whom are graduate students at Penn State with a fair knowledge of digital images and photography. All the photos used in the study are from photo.net, the same collection used in experiments. The detailed design and the evaluation of the user study are reported below.

A collection of around 1000 images were randomly picked to form the dataset for the study. Each participant is provided with a set of 160 randomly chosen images and is asked to describe the composition layout of each image. At an online site, the participants can view pages of test images, beside each of which are selection buttons for seven composition categories: “Horizontal”, “Vertical”, “Centered”, “Diagonal (upper left, bottom right)”, “Diagonal (upper right, bottom left)”, “Patterned”, and “None of Above”. Multiple choices are allowed. “Patterned” is used for the class of photos with homogeneous texture (the so-called “Textured” class). And the “None of Above” category is added to allow more
Table 5.2: Distribution of the entropy for the votes of users. For each composition category, the percentage of photos yielding a value of entropy in any bin is shown.

<table>
<thead>
<tr>
<th>Composition</th>
<th>(0, 0.5]</th>
<th>(0.5, 1.0]</th>
<th>(1.0, 1.5]</th>
<th>(1.5, 2.0]</th>
<th>(2.0, 2.5]</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>36.12</td>
<td>29.96</td>
<td>17.18</td>
<td>15.42</td>
<td>1.32</td>
</tr>
<tr>
<td>v</td>
<td>12.98</td>
<td>45.67</td>
<td>19.71</td>
<td>20.19</td>
<td>1.44</td>
</tr>
<tr>
<td>c</td>
<td>25.36</td>
<td>45.48</td>
<td>13.12</td>
<td>14.87</td>
<td>1.17</td>
</tr>
<tr>
<td>ulbr</td>
<td>12.99</td>
<td>44.16</td>
<td>19.48</td>
<td>19.48</td>
<td>3.90</td>
</tr>
<tr>
<td>urbl</td>
<td>16.87</td>
<td>43.37</td>
<td>18.07</td>
<td>20.48</td>
<td>1.20</td>
</tr>
<tr>
<td>t</td>
<td>10.77</td>
<td>36.92</td>
<td>10.77</td>
<td>36.92</td>
<td>4.62</td>
</tr>
<tr>
<td>none</td>
<td>6.59</td>
<td>39.56</td>
<td>17.58</td>
<td>34.07</td>
<td>2.20</td>
</tr>
</tbody>
</table>

flexibility for the user perception. At the end, 924 images voted by three or more users form the dataset for evaluation.

5.4.4.1 Variation in Users’ Choices of Composition

The variation in users’ votes on composition layout is examined for better understanding of compositional clarity. The ambiguity in the choices of composition layout is quantified using entropy. The larger the entropy in the votes, the higher is the ambiguity in the compositional layout of the image. The entropy is calculated by the formula \( \sum p_i \log \frac{1}{p_i} \), where \( p_i, i = 0, ..., 6 \), is the percentage of votes for each category. The entropy was calculated for all 924 photos and its value was found to range between 0 and 2.5. The range of entropy is divided into five bins for analysis (Table 5.2). The photos are divided into seven groups according to the composition category receiving the most votes. In each category, the proportion of photos yielding a value of entropy belonging to any of the five bins is computed. These proportions are reported in Table 5.2. The results indicate that among the seven categories, “Horizontal” and “Centered” have the strongest consensus among users, while “None of above” is the most ambiguous category.

5.4.4.2 Composition Classification Results

The proposed composition classification method is evaluated in the case of both exclusive classification and non-exclusive classification. The users’ votes on composition are used to form ground truth, with specifics to be explained shortly. Six categories are considered, i.e. “Horizontal”, “Vertical”, “Centered”, “Diago-
Table 5.3: Confusion table for exclusive classification of 494 images into six composition categories. Each row corresponds to a ground truth class.

<table>
<thead>
<tr>
<th></th>
<th>h</th>
<th>v</th>
<th>c</th>
<th>ulbr</th>
<th>urbl</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>107</td>
<td>0</td>
<td>20</td>
<td>3</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>v</td>
<td>1</td>
<td>32</td>
<td>39</td>
<td>3</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>c</td>
<td>10</td>
<td>7</td>
<td>132</td>
<td>8</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>ulbr</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>18</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>urbl</td>
<td>2</td>
<td>1</td>
<td>13</td>
<td>0</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>t</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>

The confusion table (Table 5.3) shows that the most confusing category pairs are “Vertical” vs “Centered” and “Diagonal_vulbr” vs “Centered”. Figure 5.10(a) shows some examples labeled “Vertical” by users while classified as “Centered” by

The exclusive classification is conducted only on photos of little ambiguity according to users’ choices of composition. The number of votes a category can receive ranges from zero to five. To be included in this analysis, a photo has to receive three or more votes for one category (that is, the ground-truth category) and no more than one vote for any other category. With this constraint, 494 images out of the 924 are selected. Table 5.3 is the confusion table based on this set of photos.

The exclusive classification is conducted only on photos of little ambiguity according to users’ choices of composition. The number of votes a category can receive ranges from zero to five. To be included in this analysis, a photo has to receive three or more votes for one category (that is, the ground-truth category) and no more than one vote for any other category. With this constraint, 494 images out of the 924 are selected. Table 5.3 is the confusion table based on this set of photos.

The confusion table (Table 5.3) shows that the most confusing category pairs are “Vertical” vs “Centered” and “Diagonal_vulbr” vs “Centered”. Figure 5.10(a) shows some examples labeled “Vertical” by users while classified as “Centered” by
the algorithm. The misclassification is mainly caused by the following: 1) “Vertical” images in the training dataset cannot sufficiently represent this category; 2) users are prone to label images with vertically elongated objects “Vertical” although such images may be “Centered” in the training data; 3) the vertical elements fail to be captured by image segmentation. Figure 5.10(b) gives “Diagonal” examples mistakenly classified as “Centered”. The failure to detect diagonal elements results mainly from: 1) diagonal elements which are beyond the diagonal tolerance set by the algorithm; 2) imaginary diagonal visual paths, for example, the direction of an object’s movement.

![Figure 5.10](image)

In non-exclusive classification, the criterion for a photo being assigned to one category is less strict than in the exclusive case. A photo is labeled as a particular category if it gets two or more votes on that category. In total there are 849 out of the 924 photos with at least one category voted twice or more. The results reported below is based on these 849 photos.

The composition categorization of a photo is represented by a six-dimensional binary vector, with 1 indicating the presence of a composition type, and 0 the absence. Let $M = (m_0, ..., m_5)$ and $U = (u_0, ..., u_5)$ denote the categorization vector generated by the algorithm and by users respectively. The value $m_0$ is set
to 1 if and only if there are 10 or more nearest neighbors (among 30) labeled as “Horizontal”. The values of $m_1$ and $m_2$, corresponding to the “Vertical” and “Centered” categories, are set similarly. For the diagonal categories, $m_i$, where $i = 3, 4$, is set to 1 if any diagonal element is detected by the diagonal element detector. Finally, $m_5$ is set to 1 if the “Textured” versus “Non-textured” classifier labels the image “Textured”. Three ratios are computed to assess the accuracy of the non-exclusive classification.

- Ratio of partial detection $r_1$: the percentage of photos for which at least one of the user labeled categories is declared by the algorithm. Based on the 849 photos, $r_1 = 80.31\%$.

- Detection ratio $r_2$: the percentage of photos for which all the user labeled categories are captured by the algorithm. Define $M \succ U$ if $m_j \geq u_j$ for any $j \in [0, 5]$. So $r_2$ is the percentage of images for which $M \succ U$. Based on the 849 photos, $r_2 = 66.00\%$.

- Ratio of perfect match $r_3$: the percentage of photos for which $M = U$. Based on the 849 photos, $r_3 = 33.11\%$.

![Figure 5.11: Histogram of the number of mismatches between classification by our algorithm and by the users.](image)

The precision of the algorithm can also be measured by the number of mismatches between $M$ and $U$. A mismatch occurs if $m_j \neq u_j$, $j = 0, \ldots, 5$. The
number of mismatches for an image ranges from 0 to 6. This number is counted for every image and the histogram of these counts is plotted in Figure 5.11.

5.5 OSCAR System

![Prototype of the OSCAR system](image)

The composition layout classifiers and composition sensitive image retrieval method presented in this Chapter are important parts of a recently proposed photography feedback system named OSCAR (On-Site Composition and Aesthetics Feedback through Exemplars for Photographers) [100]. The OSCAR system has been designed to provide on-site composition and aesthetics feedback through retrieved examples. The prototype of the system is illustrated in Figure 5.12. It includes three novel interactive feedback components, namely a composition feedback module, a color combination feedback module, and an overall aesthetics feedback module. The composition feedback module analyzes the spatial layout properties of the query image and retrieves highly aesthetic exemplar images similar in terms of content and composition from the database. The color combination feedback module estimates the confidence level of an image containing aesthetically pleasing color combinations. The aesthetics feedback module predicts general aesthetics scores for both color and monochromatic images. This system was designed keep-
ing the next generation photography needs in mind and is the first of its kind. Figure 5.13 shows the homepage of the demo website for the OSCAR system. Figure 5.14 presents a potential visual interface in mobile application.

Figure 5.13: Homepage of OSCAR demo website.

Figure 5.14: A potential visual interface for the proposed photography feedback system in mobile applications.
5.6 Summary

A novel spatial composition categorization algorithm is presented in this chapter. The composition analyzer has been integrated with an image retrieval system SIM-PLICITY and an existing aesthetics assessment module to return highly aesthetic exemplar images from the corpus which are similar in content and composition. An comprehensive system to enhance the aesthetic quality of the photographs, named OSCAR, has been designed to provide on-site composition and aesthetics feedback through retrieved examples. This system was designed keeping the next generation photography needs in mind and is the first of its kind. The feedback rendered is guiding and intuitive in nature. It is computed in situ while requiring minimal input from the user.
Chapter 6

Digital Dodging and Burning for Photographs Inspired by Notan Design of Exemplars

6.1 Introduction

Photography has been democratized due to the prevalence of digital cameras and camera phones. Hundreds of millions of snapshot photos are produced and shared every day, creating a demand for convenient photo enhancement applications. Image processing techniques have long offered ways to improve the visual quality of photographs. Filters were developed to enhance photo contrasts, reduce image noises, and create effects like a vintage style and a sketchy look. How to utilize the expertise of artists and professional photographers to help amateurs make more aesthetically pleasing photos remains an open research question. Bridging the gap between the human visual system and machine vision regarding aesthetics is a growing area of research.

The arrangement of dark and light is a basic building block in artistic design. In painting, this contrast is called Notan, following a Japanese design concept that means dark-light harmony, which is often the essential factor of an artwork’s success [70]. The study of Notan looks into the relationship between dark and light values. Mass Notan study, a key method in the field, focuses on the organization of
simplified tonal structure rather than details. For example, a scene is reduced to an arrangement of major shapes (mass) with different levels of intensities. The goal of a mass Notan study is to create a harmonious and balanced “big picture.” The same concept is also emphasized in photography. Due to the difficulty in controlling light, especially in outdoor environments, photographers use dodging and burning techniques to achieve desired exposures for regions that cannot be reached by a single shoot. Traditionally, dodging and burning are darkroom techniques applied during the film-to-paper printing process to alter the exposure of certain areas without affecting other parts of the same photo. Specifically, dodging decreases the exposure where the photographer wants more light, and burning refers to the technique that darkens particular areas. Ansel Adams extensively used dodging-and-burning techniques in developing many of his famous prints. He mentioned in his book *The Print* [6] that most of his prints are not the reproduction of the scenes but instead his visualization of the scenes. As Ansel Adams put it, “dodging and burning are steps to take care of mistakes God made in establishing tonal relationships.”

In the digital era, photographers can restyle the tonal structure with the help of photo editing software to realize their visualization of the scenes. However, applying dodging and burning can be time-consuming and involves some level of expertise in both photography and the software application. The ability to visualize scenes and create aesthetically pleasing designs also requires experience. This chapter discusses the study of dark-light composition in photographs. The problem of creating a good design is formulated from a different point of view: given an exemplar photograph with a highly-regarded design of dark and light, how can we apply its dark-light composition to another image. The dark-light composition of an image is described by its overall intensity distribution and the proportion of dark and light regions. An approach is proposed to adjust the dark-light arrangement of a photograph automatically by referring to the exemplar’s dark-light composition.

Arguably, many professional visual artists, photographers, and painters believe that the luminance channel is more important than color. A well-balanced and aesthetically pleasing luminance map is not only essential for the success of black-and-white photographs but also is critical for making good color photographs.
The proposed method focuses on manipulation of the luminance channel. We attempt to achieve aesthetic enhancement on images by adjusting their dark-light arrangements. The proposed approach is different from the existing work in several ways:

- First, instead of deriving a global tone mapping function from two intensity distributions, we implement histogram matching through region-wise adjustments. The rules of adjustments are learned from an exemplar photograph. Inspired by the dodging and burning techniques in photography, the brightening and darkening effects are added on with a region-wise touch.

- The region-wise adjustments are controlled by transformation functions. A special case of the generalized logistic function is employed as the form of transformation functions. Parameters for individual functions are estimated by optimizing an objective term.

- Inspired by mass Notan study in painting, we consider the dark-light composition of the image during the histogram-matching process. The dark-light composition of an image is interpreted by two components, the overall intensity distribution and the two-value Notan structure.

### 6.2 The Algorithm

Studying the Notan structure (the dark-and-light arrangement) of a scene is an important step in imagery art such as painting and photography. An artist may make multiple sketches and a photographer may take several trial shots to arrive at a promising design. It is extremely difficult (if not impossible) for machines to simulate art creation behavior. To compensate for this difference, we attempt to arrange the dark and light in an image by utilizing the Notan pattern of an exemplar and constraining the scope of modification to reach an acceptable design. The analysis is focuses on the luminance channel.
6.2.1 Region-wise Histogram Representation

A straightforward approach to characterizing the dark-light composition of an image is by its intensity histogram. By using histogram matching, we can compare the dark-light rearrangement of the source image and the exemplar. Because we intend to process the source image in a region-wise fashion, we first represent the overall histogram by summing up its sub-histograms, the intensity histogram of a region. The image segmentation algorithm described in [47] is used to divide an image into semantically meaningful regions. The image is converted into CIELab color space, and the luminance channel is extracted to build the sub-histograms according to the segmentation results. The range of intensity values is $[0, 1]$ throughout this paper. In this way, the intensity distribution of the entire image can be represented by

$$H(x) = \sum_{i=1}^{n} H_i(x) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \frac{1}{\sqrt{2\pi} \sigma_{ij}} \exp \left( \frac{(x - u_{ij})^2}{2\sigma_{ij}^2} \right),$$

where $H_i$ is the sub-histogram for the $i$th region and $n$ is the number of regions. Because regions formed by the image segmentation algorithm are clusters of pixels with very similar intensity values, we assume a sub-histogram is subject to either a single-mode or a bimodal distribution. The bimodal option is provided to accommodate intensity distributions of very textured regions. A sub-histogram is then approximated by a Gaussian Mixture Model (GMM) with $K_i \in \{1, 2\}$ components. An unsupervised clustering algorithm [15] is used to estimate $K_i$ as well as the mean and the variance of each component, $\mu_{ij}$ and $\sigma_{ij}$. Similarly, the intensity distribution of the exemplar $H^{(t)}$ is also approximated by GMM. Instead of summing over sub-histograms, a single GMM with $K_t$ components is used to represent the entire image. The value of $K_t$ is also estimated by the algorithm in [15]. The distance between the two intensity distributions is defined as the integrated difference between their cumulative density functions [54],

$$D(H, H^{(t)}) = \int_{0}^{1} \left( \int_{0}^{\lambda} H(x) dx - \int_{0}^{\lambda} H_t(x) dx \right)^2 d\lambda. \quad (6.1)$$
6.2.2 Transformation Functions

The region-wise intensity adjustments are implemented by transformation functions. A special case of the generalized logistic function is adopted as the form of tone mapping functions. The generalized logistic function is defined as

\[ Y(x) = A + \frac{K - A}{(1 + Qe^{-B(x-M)})^{1/v}}. \]

The general form provides a high degree of freedom. Only two parameters \(b\) and \(m\) are retained to allow changes in curvature and translation of the inflection point [64].

\[ Y(x) = \frac{1}{1 + e^{-b(x-m)}}. \]  

(6.2)

The above function was selected because it can accomplish different types of tone adjustment tasks by fitting appropriate parameters and keeping the transformation functions in a consistent form. Moreover, the logistic curve tends to preserve contrast. Figure 6.1 illustrates some tone mapping curves generated by (6.2) with different values of \(b\) and \(m\).

![Tone mapping curves](image)

Figure 6.1: Tone mapping curves with various parameters.

We constrain the parameter space such that firstly, the parametrized instance of equation (6.2) is monotonically ascending, and secondly, the range after transformation is not compressed too much. The first condition can be met provided
For the second condition, we set two thresholds $t_0$ and $t_1$ such that:

$$Y(0) = \frac{1}{1 + e^{bm}} \leq t_0, \quad Y(1) = \frac{1}{1 + e^{-b(1-m)}} \geq t_1.$$ (6.3)

Then two parameters $m_i$ and $b_i$ determine the intensity adjustment for the $i$th region. A right (left) translation of the inflection point, i.e. $m \gg 0.5$ ($m \ll 0.5$), will darken (brighten) the region, causing a burning (dodging) effect. The image-level operation can then be represented by $T = \{m_1, b_1, ..., m_n, b_n\}$. After we apply the transformation functions to individual regions, the histogram of the modified image can be approximately calculated as:

$$H(y; T) = \sum_{i=1}^{n} \frac{dX_i(y)}{dy} H_i(X_i(y); T)$$

$$X_i(y) = Y_i^{-1}(y).$$ (6.4)

The major problem with global tone mapping functions is that the two-dimensional intensity distribution is reduced to one dimension where the spatial information is completely lost. On the other hand, color transfer between matched regions is intuitive but requires correspondence between regions, which is straightforward for content-wise similar image pairs, but not always obvious for general images. For example, Figure 6.2 shows a pair of images taken as the source image and the exemplar image. Their intensity distributions are very different from each other. Figure 6.3 compares two global approaches, global histogram matching and color normalization [71], with the proposed region-wise approach. When the source image is low-keyed and the target is high-keyed, a global mapping function tends to wash out the dark areas and overexpose the light areas. With region-wise adjustments, however, each transformation function contributes to the overall histogram matching while its transformed range is not constrained by other regions. For example, tone mapping curve of a dark region can have a higher growth rate than light regions can (Figure 6.3 (a)).
6.2.3 Parameter Estimation

We cast the problem of dark-light rearrangement as an optimization problem. An objective function is defined to measure the degree of matching between the source image and the exemplar. The transformation parameters for individual regions are estimated by optimizing the objective function. Let $F(T)$ denote the objective function where $T$ represents parameters for the transformation functions. Both the source image and the exemplar are represented by the average intensities of their regions, $I = \{(\mu_1, p_1), ..., (\mu_m, p_m)\}$, $I^{(t)} = \{(\mu'_1, p'_1), ..., (\mu'_n, p'_n)\}$ where $\mu_i$ and $p_i$ are the average intensity and the area percentage for the $i$th region in image $I$ ($\mu'_j$ and $p'_j$ for the $j$th region in image $I^{(t)}$), $1 \leq i \leq m$ ($1 \leq j \leq n$). Then the objective function for region-wise adjustments can be defined as:

$$F(T) = \min_T D(H(y; T), H^{(t)}(y)), $$

s.t. $(\mu_i - \mu_j)(\mu'_i - \mu'_j) \geq 0$, 

for any $1 \leq i \leq m$, $1 \leq j \leq n$ . \hspace{1cm} (6.5)

where $D$ is the matching distance given in (6.1), $H$ is the transformed distribution calculated by (6.4), and $H^{(t)}$ is the target intensity distribution. The optimization is constrained so that the original order of region intensities is retained. The parameters are initialized using a greedy approach. Regions in both $I$ and $I^{(t)}$ are ordered by $\mu$, i.e., $\mu_1 \leq \mu_2 \leq ... \leq \mu_m$ and $\mu'_1 \leq \mu'_2 \leq ... \leq \mu'_n$. The estimated intensity for the first region is given by:

$$\bar{\mu}_1 = \frac{1}{p_1} \sum_{i=1}^{k-1} \mu'_i p'_i + \mu'_k \left( p'_k - (p_1 - \sum_{i=1}^{k-1} p'_i) \right),$$
(a) Tone mapping curves. Left: histogram matching and color normalization. Right: Transformation functions for different regions (red curve for black region; green for gray region; and blue for white region). Region division is shown below.

(b) Left to right: modified image in the order of histogram matching, color normalization and region-wise adjustment, and image regions.

(c) Histogram evolution for histogram matching. Left to right: black region, gray region, white region, and entire image. Histogram in gray represents the original; red represents the modified; and blue represents the target.

(d) Histogram evolution for color normalization.

(e) Histogram evolution for region-wise adjustment.

Figure 6.3: Comparison between global and region-wise adjustments.

\[ k = \arg \min_j \sum_{i=1}^{j} p_i' \geq p_1. \]  

(6.6)

The percentages, \( p_1, p_1', \ldots, p_{k-1}' \) and \( p_k' \) (partially), are deducted from \( I \) and \( I^{(t)} \). This continues until we obtain estimated intensities for all regions in \( I \). Note that the greedy approach guarantees the constraints in (6.5) are met. Given the constraints...
for the transformation functions in (6.3), a hypothetical estimated intensity for the
ith region as $\tilde{\mu}_i$ and the original average intensity $\mu_i$, then we can initialize $b_i$ and
$m_i$ by:

$$
\begin{align*}
\hat{b}_i &= \max \left( \log \left( \frac{1-\tilde{\mu}_i}{\mu_i} \right), \log \left( \frac{1-\tilde{\mu}_i}{1-\mu_i} \right) \right), \\
\hat{m}_i &= \mu_i + \frac{1}{\hat{b}_i} \log \frac{1-\tilde{\mu}_i}{\mu_i}.
\end{align*}
$$

(6.7)

6.2.4 The Impact of Two-value Notan

However, the objective function provided in (6.5) has little control on the overall
contrast of the modified image. The matching does not consider the spatial ar-
rangement of the dark and the light or the contrast structure present in the image.
When the dark-light proportions are very different in the source image and the ex-
emplar, then matching is likely to dilute the overall contrast. For example, because
the exemplar image in Figure 6.4 (b) has a small portion of dark areas (rocks)
that contrast with the large light areas, while the source image has a relatively
large dark proportion, the dark-light contrast of the modified image in Figure 6.4
(c) is diluted after the transformation. In order to retain contrast in the modified
image, we introduce into $F(T)$ the two-value Notan factor. Both the source image
and the exemplar are reduced to their two-value Notan structures. We first obtain
a binarization threshold using Otsu’s method [61] which assumes a bimodal distri-
bution and calculates the optimum threshold such that the two classes separated
by the threshold have minimal intra-class variance. Its average intensity and the
binarization threshold determine the dark-light property of a region. The distance
between the two intensity distributions is calculated for the dark areas and the
light areas separately. Matching the dark and the light separately tends to pre-
serve the overall contrast and avoid flattening the image. $H_{\text{dark/light}}$ and $H_{\text{dark/light}}^{(t)}$
in (6.8) are normalized first. Figure 6.4 (d) shows a modified image generated
by (6.8) which has less dilution in the dark area.

$$
F_1(T) = \min_T \left( D(H_{\text{dark}}, H_{\text{dark}}^{(t)}) + D(H_{\text{light}}, H_{\text{light}}^{(t)}) \right), \\
s.t. (\mu_i - \mu_j)(\mu_i' - \mu_j') \geq 0, \text{ for any } 1 \leq i \leq m, 1 \leq j \leq n.
$$

(6.8)
However, when the original two-value Notan structure is not desirable, we can modify the dark-light composition by imposing on the source image a new Notan structure in the histogram matching stage so that the dark-light composition of the modified image is more balanced. Figure 6.5 illustrates the impact of Notan structure on the modification process. By imposing different two-value Notan structures, the modified images generated by (6.8) present very different dark-light composition. Two different Notan structures are shown for each scene in Figure 6.5. The first two rows include modified images and imposed Notan structures. The same exemplar is used to generate the two modified versions. The last two rows show a source image and its original Notan structure. The source image itself is taken as the exemplar, and the modified version is generated by imposing a different Notan structure on the source image. The results clearly demonstrate the power of the underlying Notan structure, which means that even without a good exemplar, it is possible to boost a better picture just by adjusting its Notan structure.
Figure 6.5: Modification by different Notan structures.


6.2.5 Optimization of Notan for Source Image

Usually images with a balanced Notan structure give a more aesthetically pleasing look. We assume that the exemplar always has a well-balanced Notan structure and then we derive a new Notan structure for the source image based on the exemplar. The Notan structure imposed on the source image is generated in such a way that each region in light areas has an average intensity higher than any region in dark areas. By relaxing the binarization threshold, we achieve a new Notan structure which has a dark-light proportion closest to the exemplar’s. The exemplar image is not only used to adjust the intensity distribution, but also it is used to determine the Notan design of the modified image. It is also possible to specify the Notan structure by manually putting image regions to the dark (light). The Notan structures in Figure 6.5 are created in this way based on the image segmentation results. For experiment results in the rest of the paper, the automatic setting is employed.

6.2.6 Compositing Modified Regions

Compared with traditional histogram manipulation algorithms, one advantage of applying transformation functions in a region-wise fashion is that it yields smooth regions without introducing noisy artifacts. However, its performance depends on region segmentation to some extent. If the same object is mistakenly segmented into several regions, different transformation functions applied on its parts can cause artifacts. In real dodging and burning practice, a similar situation can be remedied by a careful localized motion of the covering material during the darkroom exposure development or by applying a subtle dodging/burning brush over a large area in digital photo editing software. We use fuzzy region maps to cope with this problem. A bilateral filter is employed to generate fuzzy maps for regions. Bilateral filter is well known for its edge-preserving property. It considers both spatial adjacency and intensity similarity. We use the fast implementation in [63]. The fuzzy map for the \( i \)th region is defined by:

\[
m_p = \frac{1}{k_p} \sum_q G_{\sigma_s}(\|p - q\|)G_{\sigma_r}(\|I_p - I_q\|)l_q,
\]
\[ k_p = \sum_q G_{\sigma_s}(\|p - q\|)G_{\sigma_r}(\|I_p - I_q\|). \] (6.9)

\[ l_q \] is a binary label for the pixel at \( q \), i.e. \( l_q = 1 \) if pixel \( q \) belongs to the \( i \)th region and 0 otherwise. The sizes of the spatial kernel and the range kernel are set to be 25 and 0.1 respectively. Figure 6.6 shows a fuzzy region map example.

Figure 6.6: Fuzzy region map. From left to right: original image, image segment and its fuzzy map.

Suppose the source image is divided into \( n \) regions, and their fuzzy maps are \( \{m_i, i = 1, ..., n\} \), and \( \sum_{x,y} m_i(x, y) = s_i \) (\( s_i \) is the size of the \( i \)th region). We then modify the image according to (6.10), where \( l_i(x, y) \) is the new intensity value at \( (x, y) \) after applying the transformation function for the \( i \)th region. The fuzzy region maps work well for most cases, but gradient reversal may occur along the region boundary when a region’s transformation function has a steep curvature (second example in Figure 6.9).

\[ l'(x, y) = \sum_{i=1}^{n} m_i(x, y) \times l_i(x, y). \] (6.10)

### 6.3 Automatic Dodging and Burning Results

In Section 6.2, we discussed approaches to quantifying the difference between images’ dark-light arrangements. Our goal is to aesthetically enhance the source image by minimizing the difference between its dark-light composition and the exemplar’s. Features we use to characterize the dark-light composition include the intensity distribution and dark-light proportion. Figure 6.7 shows two examples of using dodging and burning in Photoshop. We attempt to achieve a similar visual
effects by transferring the dark-light arrangement of an exemplar. The first column of Figure 6.8 presents exemplar photographs with different dark-light compositions. We use the objective function $F_1$ in (6.8) to suggest transformation parameters for different regions. Figure 6.8 compares modified images output by our algorithm, global histogram matching and color normalization [71]. Because of the region-wise adjustments and the regulation of parameter space, the proposed method tends to generate smoother histograms, and the modifications made on the original images are well controlled.

Figure 6.7: (a) The source image, and (b) a modified version in Photoshop.

Figure 6.9 gives more examples of applying the automatic dodging and burning approach. From the results, we observe that the global histogram matching often yields artificial sudden changes in intensity. The color normalization method uses a linear mapping function whose growth rate is determined by the variances of the source and the target distributions. A high (low) growth rate can burn out (flatten) the resultant image. The proposed approach provides better control on extreme cases by regulating the transformation parameters. One problem with the proposed method is that the intensity adjustments depend on the image segmentation results to some extent. However, if there is no severe segmentation error, the algorithm can output reasonable results with some level of aesthetics enhancement. In the experiment setting, the number of segments is set to be 3 for simple scenes and 6 for complex scenes. Note that more segments correspond to more parameters to be estimated and therefore more computation.
Figure 6.8: Modified images and their histograms. From left to right: exemplar photographs, output by global histogram matching, color normalization and the proposed algorithm.
Figure 6.9: More experiment results. (a) the source image; (b) the exemplar; (c) results by global histogram matching; (d) results by color normalization [71]; (e) results by the proposed approach.
6.4 Summary

An approach to emulate the dodging and burning techniques in photography is presented in this chapter. Dodging and burning are darkroom photographic printing techniques for replicating the tonal relationship of the high range dynamics of the real world in a low range media. They also offer opportunities to enhance the aesthetics of photographs, especially in black-and-white photography. While existing photo editing software can perform dodging and burning, the process often involves tedious adjustments. Inspired by concepts like Notan in art and the darkroom dodging and burning process, the proposed method improves the dark-
light composition by adjusting exposures in a region-wise fashion. An exemplar photograph is used to control the overall look of the modified image. Experimental results show that the proposed approach can effectively transfer the illumination of the exemplar while considering the aesthetics constraints.
Chapter 7

Conclusion and Future Work

This dissertation studies two novel problems under the umbrella of computational aesthetics, specifically characterization of painting style and analysis of photography composition. The goal is to contribute to the newly emerging discipline of computational aesthetics and to explore new approaches to enhancing user experience in activities related to aesthetic interpretation of images.

With regard to characterization of painting style, we have shown that painting styles of particular artists can be characterized using numerical features derived from automatically extracted brushstrokes. The state-of-the-art research on digital painting analysis mainly focus on profiling low-level features. Extraction of real brushstrokes have not been well studied prior to this work. The proposed algorithms have demonstrated success on tackling real-world painting analysis tasks. The results show great possibility of exploring computational methods as new scientific tools for art historians.

For analysis of photography composition, we have presented algorithms to classify general photographs into different spatial composition categories. Specifically, we designed a diagonal element detector according to principles in photography and psychology, and a new spatial signature to characterize and then distinguish the spatial layouts of photographs. An image retrieval system can be enhanced with spatial composition categorization by retrieving photographs with similar composition. The composition sensitive retrieval module has been further integrated with an aesthetics assessment module and a color evaluation module to render on-site feedback for photographers. Existing works on automatic composition enhance-
ment are mostly offline in nature and therefore limit the scope of improvement. The proposed interactive system aims to provide amateur photographers with new experience in learning photography.

We also studied photography composition from the perspective of dark-light arrangement. Traditional tone reproduction and color transfer methods mainly focus on boosting low-level image quality without considering high-level aesthetic constraints. We propose a region-wise dark-light rearrangement algorithm by utilizing the Notan structure and intensity distribution of an exemplar. The new approach emulates the dodging and burning techniques in darkroom photography to enhance photograph aesthetics while avoiding undesirable artifacts by restraining the modification process.

7.1 Future Work

Similar to semantic analysis, computational studies on aesthetics attempt to duplicate human perception of aesthetics on machine. This dissertation made several novel contributions to the interpretation of aesthetic terms for two topics, but there is much room for improvements and extensions.

The proposed algorithms for brushstroke analysis and curve elegance measurement have been applied on paintings of van Gogh and Norval Morrisseau respectively. The scope of application of these algorithms can be further explored in different painting analysis tasks. Brushstroke is the most primitive element artists put on the canvas. Other high-level visual elements and structures can also be consiered as the main subjects for style characterization. Current results show great possibility for in-depth analysis of oil paintings through computational approaches.

The proposed photography feedback system is the first known attempt to automatically generate on-site integrated photography feedback for photographers. Because of the inherent complexity of aesthetics and the multifaceted nature of picture composition, there is ample room for extension and enhancement in the future. The composition categorization can be refined further to include more classes. A significant step beyond the present analysis of composition is to make on-site automatic suggestions about placement of objects in a photo, which can be
achieved by zooming, expanding, tilting, etc. Principles of good composition can be applied based on composition characteristics extracted by the computer. For instance, the diagonal spatial structure provides a sense of dynamism in the image and is highly pleasing. Adjustment made to the image frame to diagonalize the not-so diagonal element can increase the image aesthetics.
Bibliography


Vita
Lei Yao

Lei Yao was born in Hangzhou—the capital city of Zhejiang province, China—on September, 1983. She is the only child of Caiping Qu and Xinhao Yao. After completing her degree at Zhejiang Xiaoshan High School, in 2001, she entered Zhejiang University, receiving the degree of Bachelor of Engineering from Mixed class, Chu Kochen Honors College in June 2005. Between 2005 and 2007, she studied Computer Science in the College of Computer Science at Zhejiang University, receiving a Master of Science degree in June 2007. She entered the doctoral program in Information Sciences and Technology at Penn State in August 2007, and received a Ph.D. in August 2013. Her research interests lie primarily in the general area of image processing and computer vision, with a special interest in computational photography, image retrieval, aesthetics inference and painting analysis.