MODELING PHOTOGRAPHIC COMPOSITION
VIA TRIANGLES

A Thesis in
Information Sciences and Technology

by
Siqiong He

© 2015 Siqiong He

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2015
The thesis of Siqiong He was reviewed and approved* by the following:

James Z. Wang  
Professor of Information Sciences and Technology  
Thesis Adviser

C. Lee Giles  
David Reese Professor of Information Sciences and Technology

Jia Li  
Professor of Statistics

Zihan Zhou  
Research Associate of Information Sciences and Technology

Carleen Maitland  
Associate Professor of Information Sciences and Technology  
Affiliate Professor of International Affairs  
Interim Associate Dean of Information Sciences and Technology

*Signatures are on file in the Graduate School.
Abstract

Modeling photographic composition is important to many real-world applications such as digital photography, image retrieval, image understanding, and image aesthetics assessment. The triangle technique is among the indispensable composition methods which professional photographers often rely on. It refers to the organization of visual elements such as lines, edges, and shapes into triangles, resulting in interesting and dynamic compositions. This thesis proposes a system that can detect the presence of the triangle composition in two major categories of photographs: natural scenes and portraits. After the detection, the system further provides on-site composition feedback to the photographers. In the case of natural scene photography, we propose a new image segmentation algorithm for extracting triangles based on agglomerative clustering, utilizing both photographic and geometric cues. We further illustrate how these cues can be directly used to detect the dominant vanishing point in an image without extracting any line segments. For portrait photography, we extract and assemble straight lines to form triangles via a modified RANSAC algorithm. Experimental results have demonstrated that our system can accurately locate preeminent triangles in photographs without any knowledge about the camera parameters or lens choices. Finally, we demonstrate an application of the proposed techniques in providing on-site feedback to photographers.
# Table of Contents

| Page |
|------|
| List of Figures | vi |
| Acknowledgments | ix |
| **Chapter 1. Introduction** | 1 |
| 1.1 Aesthetic Experience in Psychology | 4 |
| 1.2 Triangles in Photographic Composition | 9 |
| 1.3 System Overview | 11 |
| 1.4 “Pie” Model in Natural Scene Photography | 13 |
| 1.5 Modeling Composition in Portrait Photography | 17 |
| 1.6 Contributions | 19 |
| 1.7 Outline of the Thesis | 20 |
| **Chapter 2. Related Work** | 21 |
| 2.1 Photo Aesthetics | 21 |
| 2.2 Composition Modeling | 23 |
| 2.3 3D Modeling and Segmentation of Natural Images | 24 |
| 2.4 Portrait Photo Analysis | 25 |
| **Chapter 3. Modeling Composition in Natural Scene Photography** | 27 |
| 3.1 Hierarchical Image Segmentation | 27 |
| 3.2 Geometric Distance Measure | 29 |
## List of Figures

1.1 The use of triangles in visual art and architectural works. The suggested triangles are indicated in red. From left to right: **(Row 1)** Afghan Girl, Eiffel Tower, Girl with a Pearl Earring, Golden Gate Bridge, Mona Lisa, The Arnolfini Portrait, The Scream. **(Row 2)** Iwo Jima Memorial, Last Supper, Pyramid, Starry Night Over the Rhone, The Birth of Venus. 10

1.2 Overview of our composition analysis system. 12

1.3 Applying pie models on natural scene photos. 14

1.4 Geometric image segmentation. **(a)** The original image with the dominant vanishing point detected by our method (shown as a yellow round dot). **(b)** Region segmentation map produced using a state-of-the-art method. **(c)** Geometric image segmentation map produced by our method. 15

3.1 Hierarchical image segmentation using photometric cues only. **(a)** The original image. **(b)** The ultrametric contour map (UCM) generated by Arbelaez et al. [2]. **(c)** The segmentation result obtained by thresholding the UCM at a fixed scale. 29

3.2 Illustration of the computation of the geometric distance. **(a)** The oversegmentation map with the polar coordinate system. **(b)** Three adjacent regions from the image. **(c)** The histograms of angle values for the three regions. **(d)** The boundary map weighted by the geometric distance between adjacent regions. 32
3.3 Image segmentation results by integrating both photometric and geometric cues. Different weighting parameter $\lambda$ have been used.

3.4 Additional segmentation results. For each original image, we show the results with $\lambda = 1, 0.8, 0.6, 0.4, 0.2$ and 0, in this order respectively.

3.5 Enhancing vanishing point detection. (a) Original image. (b) Line segment detected. (c) and (d) The weighted boundary map for two different hypotheses of the dominant vanishing point location. (e) The consensus scores for all vertices on the grid.

4.1 Examples of implied triangles in portrait photography.

4.2 LSD results with different density.

4.3 Filtering the line segments. The parameter $\alpha$ is set to 0.5 and 0.7 in (d) and (e), respectively.

4.4 Challenges in detecting triangles: Outliers and imperfect sides. (a) outliers: irrelevant objects, (b) outliers: multiple triangles, (c) imperfect sides: occlusion, (d) imperfect sides: bending curve.

4.5 Illustration of the RANSAC algorithm. (a) original image, (b) candidate sides, (c) neighborhood region, and (d) fitted pixels.

5.1 Segmentation benchmarks. $K$ is the number of regions. Top rows show some example test images with the manually labeled segmentation maps.

5.2 Composition-sensitive image retrieval results. Each row shows a query image (leftmost) and the top-6 or top-8 ranked images retrieved.

5.3 Detected triangles with different Continuity Ratios
5.4 Examples of detected triangles in portrait photographs. . . . . . . . . . 58
5.5 The performance of portrait retrieval system. . . . . . . . . . . . . . 60
5.6 Useful and useless examples of retrieved portraits. Each row contains results from the same query. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 62
Acknowledgments

This thesis would not have been possible without the guidance, motivation, and help from many individuals. This is my sincere attempt to acknowledge these people.

First of all, I would like to acknowledge my thesis advisor, Prof. James Z. Wang, for his invaluable guidance and support. He has been a constant source of sparkling inspirations and insightful advices throughout my graduate study. I also want to thank Dr. Zihan Zhou who has been cooperating with me on my research project. I appreciate him for being my mentor, providing detailed guidance, critiquing my work and helping me polish my ideas. Furthermore, I acknowledge Prof. C. Lee Giles and Prof. Jia Li for serving on my thesis committee. Their thorough and excellent feedbacks give me a strong foundation to continue on my research topic all the time.

I also would like to thank all my fellow graduate students at Penn State University: Mu Qiao, Lei Yao, Neela Sawant, Yu Zhang, Xin Lu, Baris Kandemir, Jianbo Ye, and Yukun Chen. I am grateful for all their generous support and help.

Last but not least, I would like to thank my parents Qiting He and Caijuan Zhang as well as my husband Jia Zhan for their love and support during these years, without whom I would not have such an accomplishment.
Chapter 1

Introduction

With the rapid advancement of digital camera and mobile imaging technologies, we have witnessed a phenomenal increase of both professional and amateur photographs in the past decade. Large-scale social media companies, e.g., Flickr, Instagram, and Pinterest, further empowered their users with the capability to share photos with people all around the world. As millions of new photos are added to the Internet on a daily basis, there is an increasing demand for creating automatic systems to manage, assess, and edit these photos.

Currently, most existing online image management or photo sharing services heavily rely on textual information such as tags or descriptions provided by users. A few can also access visual features of photos and infer embedded semantic information from these features. Either textual or semantic information only reveals contents of photos such as present objects, events, themes, and so on. However, the explosive growth of online photos requires those automatic systems to index them with more dimensions other than contents. For example, millions of users upload photos about birthday parties everyday. As a result, searching photos with the keywords “birthday party” generates a huge number of results. Among them, only a small portion has high aesthetic quality and thus has potential appeal to other users. Therefore, an automatic evaluating technique is in
demand to filter out these low-quality photos and enable users to easily get access to high-quality photos.

The aesthetics of photography lies in several different aspects: color, lighting, texture, composition, etc. Professional photographers are skillful at adjusting photography parameters such as aperture and shutter speed to improve multiple aspects of photos. For instance, shallow depth-of-field technique is utilized to generate specific textures, i.e., blurred background and sharp foreground, by selecting a large aperture. Compared to the blurred background, the sharp foreground immediately drags viewers’ attention. Such contrast between blurred and sharp textures makes the photo more aesthetically appealing. Once a computer can recognize such contrast, it is able to evaluate aesthetic quality of this photo from the aspect of texture. Hence, some researchers begin to explore the possibility of extracting aesthetics-relevant visual features from a photo in terms of its color, lighting, and texture. Further, the relationship between such visual features and aesthetic quality of photos can be modeled via machine learning techniques. Given a new photo without any texural or semantic information, the evaluating system can predict its aesthetic quality by extracting and tossing those aesthetic-relevant visual features into the learned model.

Accurately predicting aesthetic quality of photos has a wide range of potential applications. Nowadays, the emergence of digital cameras and mobile imaging technology makes people take photos more casually, leaving a lot of low-quality photos in their personal galleries which are unlikely to be accessed in the future. Automatically picking out low-quality photos is beneficial for large-collection personal gallery management [11, 37]. Moreover, such algorithm can also be embedded into online image search engines
as a re-ranking tool to deliver high quality images to users [36, 11, 42, 37, 35]. Apart from filtering out low-quality photos, users may want to enhance the quality of these photos. In this case, image processing techniques can be adopted together with the aesthetic quality evaluating technique to enhance the quality of consumer-level photos. Additionally, aesthetic quality assessment can also be used for story illustration [11] and realtime photographic guidance [37, 65].

Compared to color, lighting, and texture, composition is more difficult to model because it requires an overall semantic comprehension of the photo. Recently, photo composition understanding is becoming a noteworthy area of research in the computer vision community.

Composition is the art of positioning or organization of objects and visual elements (such as color, texture, shape, tone, and depth) within an image or a visual art work. Known principles of organization include balance, contrast, geometry, rhythm, perspective, illumination, and viewing path. Automated understanding of photo composition has been shown to benefit several applications such as summarization of photo collections [41] and assessment of image aesthetics [42]. It can also be used to render feedback to the photographers on the aesthetics of their photos [66, 65], and to suggest improvements on the image composition through image re-targeting [33, 6]. In the literature, most studies on image composition understanding focus on image-based rules such as the simplicity of the scene, visual balance, the rule of thirds, and the use of diagonal lines. Because of their simplicity, these composition rules have been widely used to guide the photographers at the moment of their creative work.
However, these rules are quite limited in capturing the wide variations in photographic composition. As an expansion, we hereby explore methods to identify an important composition technique, namely, the triangle technique. We focus on two major categories of photographs: natural scenes and portraits. These two categories cover a number of important consumer photography subject types including landscape, architecture, travel, portrait, fashion, family, and baby photography. Other types, such as animal, flower, macro, sport, and event photography, are frequently done by more sophisticated or professional photographers and the triangle technique is not as important for them as some other techniques such as depth-of-field controlling, high speed, motion blurring, and cropping, hence are not covered in this work.

We have developed category-sensitive approaches to detect the presence of the triangle composition and provide on-site feedback to photographers. In a realistic application, users can first inform the system the type of photography being created in order to receive the feedback. This is similar to a command dial selector, often provided on consumer-level digital cameras, where photographers can set the type of photo taking with typical choices as portrait, landscape, macro, sport, and night. In the following section, we discuss the use of triangles in composition.

1.1 Aesthetic Experience in Psychology

To gain deeper understanding of an aesthetic experience, we look into the concept of aesthetics in psychology. Aesthetic experience is defined as a complex interaction among properties of art objects, characteristics of the viewers, and the physical, social,
and historical contexts [34, 23]. It intrigues researchers from both psychology and neuroscience, giving rise to two research domains related to perceptual/cognitive aesthetics: experimental aesthetics (psychology of aesthetics) and neuroaesthetics [25].

Experimental aesthetics is the second oldest branch of experimental psychology [23]. Conceptual models and frameworks proposed in this area mainly focus on three questions: What factors will influence aesthetic experience? What is the underlying neural, perceptual, and cognitive process of aesthetic experience? How does such aesthetic experience evolve/change over time with regard to evolutionary, historical, and cultural changes? The same questions are also approached in the area of neuroaesthetics. The major difference is that researchers from neuroaesthetics investigate the neural underpinnings of aesthetic experience while concerning these questions. With the aid of imaging and neurophysiological techniques such as functional magnetic resonance (fMRI), magnetoencephalography (MEG), and electroencephalography (EEG) [8], they collect evidence to explain the relationship between biological activation of the brain and psychological process of aesthetic experience [45, 7, 8, 46].

Aesthetic experience refers to the process that human beings gain aesthetic pleasures while appreciating and evaluating the beauty in a range of entities such as painting, sculpture, music, faces, flowers, food, etc. Although most of previous work focuses on studying aesthetic experience related to visual art, the results and conclusions can be easily extended to other forms of art such as music and dance. Viewer, artwork, and context are the three elements involved in the process of aesthetic experience. To understand the three elements, we can imagine a scene where a person is looking at a painting in a museum. The person is the viewer. His/her individual characteristics
such as neural faculty, memory, expertise, and personal preferences will affect the way he/she appreciates the beauty of the painting. Likewise, the content, form, and style of the artwork, which refers to the painting in the above example, also influences the aesthetic experience. Finally, the context can be specifically stated as when and where the aesthetic experience takes place. To bring it up to a more general level, factors like historical, cultural, and economic background as well as the evolutionary process can all be treated as contextual factors in the model of aesthetic experience. These diverse factors were divided into seven perspectives by Jacobsen [23]. Besides, these factors do not prevail over the entire process of aesthetic experience. Instead, it is only a combination of several of these factors that militate in different phases of aesthetic experience.

To understand how these factors affect the way people appreciate artworks, we need to explore more about the underlying perceptual and cognitive process of aesthetic experience. From a psychological perspective, aesthetic experience is regarded as an information processing model involving five stages: perceptual analyses, implicit memory integration, explicit classification, cognitive mastering and evaluation, as well as affective and emotional processing [30].

First, the artwork is analyzed perceptually. Varieties of aesthetic-specific perceptual visual features are studied. Among which contrast, visual complexity, color, symmetry, grouping and order (where balance is concerned), etc. have positive effect on evoking aesthetic pleasures in human beings. A number of neuroaesthetic researchers explained why people prefer such features from the perspective of biology. For example, the symmetric property of most biologically important objects such as predator, prey, or mate “trained” human beings to be intrigued by symmetric objects [46].
Implicit memory integration processes features through the faculty of memory. For instance, familiarity and prototypicality ("the amount to which an object is representative of a class of objects" [30]) are processed in this stage. Similarly, what happens in this stage is also evidenced by work from neuroaesthetics. Ramachandran and Hirstein [46] well interpreted Peak Shift effect* as the action to powerfully activate the same neural mechanisms in viewer’s brain by amplifying the signal.

Until now, aesthetic experience still stays on perceptual level. The third stage, namely explicit classification, takes content and style of the artwork into account, where the viewer generalizes semantic and stylish information from perceptual features abstracted before. Viewers’ expertise as well as contextual information thus come into play.

Thus, cognitive mastering and evaluation stages together build a feedback loop where the results of cognitive mastering are continuously evaluated. Such mechanism guarantees a gain of aesthetic pleasures after successful understanding of the art object. Conversely, a subjectively perceived failure of understanding the artwork triggers restarting the previous stages where perceptual visual features are extracted and analyzed again. In neuroaesthetic, researchers have stated that the feelings of pleasure are induced because the understanding of artwork activates the rewarding areas in the brain [49, 46]. It has explained that why aesthetic experience is intrinsically positive.

There are two outputs of the model: aesthetic emotion and aesthetic judgments [7, 30]. The state of aesthetic emotion keeps changing during different stages based on

* A psychological phenomenon is typically known as its application in animal discrimination learning. In the peak shift effect, animals sometimes respond more strongly to exaggerated versions of the training stimuli. Please refer to wikipedia for more information.
the fluency of information processing [49]. To state it simply, the more fluently one understands the artwork, the more pleasure he/she gains from aesthetic experience.

The last question concerns why human beings are born with the abilities to discern beauty from ugliness. Some researchers answer this question by drawing evidence from evolutionary theories. Such theories usually relate human being’s intrinsic abilities of judging aesthetics to survival demands during evolution. The example about preference for symmetric objects can be one instance. Stebbing [55] also proposed an interesting theory that our aesthetic behavior has evolved from our innate ability to recognize the diversity of organic form for foraging purpose.

According to the psychological model of artistic experience, we have to go through five stages to process all the information contained in an artistic display. During the five stages, a variety of information resources can provide clues for aesthetic judgments such as perceptual and cognitive features of the images, as well as personal preferences and individual experience, etc. However, the state-of-the-art techniques to computationally evaluate image quality only make use of perceptual features such as color [40], texture [35], brightness [24], etc. or a few low-level cognitive information such as familiarity [9] of the images. This is partly due to the lack of mature cognitive information extracting technologies in the domain of image processing. However, it also reveals a large set of potential research questions which can be explored in the future. For instance, considering the diversity of individual cognition abilities and personal preferences, computational models of aesthetic judgments can vary across different users or user groups. Therefore, personalized systems can be studied in the future. Another inspiration of new research question comes from the stage of explicit classification in aesthetic experience
where people resort to their expertise and knowledge about artistic styles to infer the aesthetic quality of images. Therefore, we can establish a database of artistic styles or templates for all kinds of visual displays such as painting, photos, and interface design. To build such a database, we can collect training data from artists by constructing an online art community. Artists can contribute their knowledge of artistic styles by correctly labeling images or patches of images with style tags. Then, computational models can be employed to learn the relationship between the style category and the perceptual features extracted from the images or the patches of images. During this process, a computer can learn to “recognize” artistic styles. Similarly, we can also explore other potential research questions concerning different stages of the artistic experience.

For this thesis, we choose to focus on the first stage. Aside from commonly-used perceptual features such as color and lighting, we aim at studying the overall comprehension of photo compositions. In the next section, we will briefly introduce the composition photography technique studied in this work.

1.2 Triangles in Photographic Composition

In pictorial art, good composition is considered as a congruity or agreement among the elements in a design [29]. The design elements appear to belong together as if there are some implicit visual connections between them. Another term to describe this form of unity is harmony. By reflecting this principle in photography, subjects in one scene should not aimlessly scatter around. Instead, they should unify to provide an overall impression to the viewer. To convey such unity in photos, professional photographers
have designed dozens of executable techniques for composition. One universal and inter-
esting technique is to embed basic geometrical shapes in photographic compositions [61].

Human beings begin to learn about basic geometrical shapes such as circles, rectangles, and triangles since very young age. Even toddlers are capable of recognizing those basic shapes immediately. Valenzuela [61] suggests that we can explicitly or implicitly em-bed basic geometrical shapes in photos to attract viewers. Moreover, since these shapes are instantly recognized, subjects bounded within such shapes or implicitly constructing such shapes are perceived as a unity.

Among all basic geometric shapes, triangle is arguably the most popular shape utilized by professional photographers to make a composition more interesting. Such compositional technique is called “the triangle technique”. We can also find numerous examples of the use of triangles in visual art and photographic works (Figure 1.1).

In our work, we aim at detecting the usage of the triangle technique in two common types of photographs: natural scenes and portraits. We propose an automated
system that can accurately locate a variety of triangles, even those that are carefully
designed by professional photographers but difficult to be recognized by amateurs. By
detecting the usage of the triangle technique, we can model various composition types
and retrieve images based on similarity or dissimilarity in composition. We can also help
amateur photographers gain deeper understanding from professional works and inspire
them to generate photographs with more interesting composition and higher aesthetic
quality.

1.3 System Overview

We now provide an overview of the proposed composition analysis system. At
first, users are asked to indicate the type of photography they are currently engaged in.
The system can provide assistance if the selected type is natural scene or portrait. Then,
users can take a test shot and upload it to our system as a query image. Given the query
image and its category, our system will analyze the composition, particularly the use of
triangles, in the image and further provide on-site feedback to users. Figure 1.2 briefly
introduces the framework of our composition analysis system.

For natural scenes, the composition of a query image will be represented by a geo-
metric segmentation of the image. The segmentation takes into account both geometric
and photometric cues to partition an image into cohesive triangular regions and describes
how these regions together construct a composition. In our system, the geometric and
photometric cues are obtained by detecting the vanishing point and extracting the con-
tour map of an image, respectively. Utilizing segmentations to represent compositions of
images enables us to retrieve images based on compositions. As a result, our system can
Is it a natural scene or portrait photo?

Natural Scene

Vanishing Point Detection & Contour Map Extraction

Segmentation Representation of Composition

Composition-based Image Retrieval Results

Portrait

Line Segment Detection & Contour Map Extraction

Filtered Line Segments based on Contour Map

Detected Triangle (Two Sides) in Portrait

Fig. 1.2 Overview of our composition analysis system.
recommend high-quality photos with similar compositions as the query image to users. Besides, we can also retrieve images with a variety of compositions and inspire amateur photographers to employ potentially different but more interesting compositions.

For portrait, potential line segments are first extracted from a query image. Then, we filter the line segments using the contour map to remove segments with low confidence scores. Finally, the system identifies all potential triangles from the image by grouping the remaining line segments. Detecting triangles in high-quality photos taken by professional photographers helps amateurs to gain deeper understanding and inspirations about the triangle techniques. Additionally, as on-site feedback to users, our system returns a variety of triangles of different sizes, shapes, and orientations in images which might be easily overlooked by amateurs.

In the following sections, we briefly introduce the major challenges of modeling compositions of natural scenes and portrait photos as well as the algorithms we developed to solve the problem.

1.4 “Pie” Model in Natural Scene Photography

A popular compositional technique in photography is to utilize perspective effects to generate an immersive and stereoscopic scene. Such compositions emphasize a sense of 3D space within 2D display. However, existing composition modeling techniques for natural scenes are mainly concerned with the 2D rendering of objects in an image [41, 42, 66, 65, 33, 6]. In order to model good compositions with high perspective effects, it is necessary to examine the 3D structure of the scene. According to the perspective camera geometry, all parallel lines in 3D space converge to a single vanishing point in a
photo, generating a set of triangular regions (Figure 1.3). In order to convey a strong impression of 3D space and depth to viewers, the vanishing point has to lie within or near the image frame and associates with the dominant structures of the scene (e.g., grounds, large walls, bridges).

Fig. 1.3 Applying pie models on natural scene photos.

Figures 1.3 and 1.4 show some examples. We regard such a vanishing point as the dominant vanishing point of a particular photo. As one adjusts shooting angle, both the location of the dominant vanishing point as well as the sizes, shapes, and orientations of triangular regions relating to the vanishing point will change. An experienced photographer often utilizes such technique to produce various image compositions that convey different messages or impressions to viewers. The resulting model looks like a “pie” that has been cut into slices. Each slice is corresponding to one triangle. Hence, we name it the pie compositional model. Such a pie model enables users to extract compositions from photos and further edit the compositions by moving the vanishing point or resizing the slices.

To extract the pie composition from a landscape photo, we need to first determine where the pie locates with respect to the photo and then “cut” the pie into both photometrically and geometrically consistent slices, e.g., walls, sky, grounds. Locating
the pie is equivalent to detecting the origin of pie, i.e., the dominant vanishing point in the photo. In order to cut the pie, we propose a hierarchical segmentation algorithm based on agglomerative clustering which employs both photometric and geometric cues. As shown in Figure 1.4(c), such a segmentation naturally provides us a holistic comprehension of the 3D spatial structure of the scene and successfully captures a variety of different compositions. Nevertheless, obtaining such a holistic representation is a challenging problem for the following reasons.

First, detecting vanishing point itself is a very challenging problem. Most state-of-the-art methods rely on extracting all straight lines from the image which converge to a same point and identify that point as a vanishing point. Such converging straight
lines often come from man-made objects such as edges of buildings. However, some landscape photos depicting natural scenes like mountains or lakes lack cues of straight lines (Figure 3.5). To tackle this problem, we propose a novel vanishing point detecting algorithm utilizing global photometric and geometric cues instead of edges detected from local gradient information. Specifically, we first select several candidate locations of dominant vanishing points from a photo and then apply our segmentation algorithm which aggregates global photometric and geometric cues within potential slices. The segmentation with maximum photometric and geometric distances between adjacent slices identifies the most confident location of the dominant vanishing point. We have shown that our algorithm succeeds in detecting vanishing points in outdoor natural scene photos where straight lines are often absent (Figure 3.5).

Second, once the dominant vanishing point is detected, segmenting this pie into geometrically consistent planes is also not trivial. There are a number of examples where adjacent planes do not differ significantly in color or texture (e.g., the walls and the ceiling in the second picture of Figure 1.4). However, existing image segmentation algorithms primarily employ photometric cues to calculate distance between adjacent regions (see Figure 1.4(b) for examples). To overcome this problem, we propose a geometric distance to describe the locational relationship of two regions with respect to the dominant vanishing point. Specifically, consider a polar coordinate system whose origin locates at the dominant vanishing point in the image, a region can be represented as a set of points that cover a specific range of angles. The geometric distance measures how two regions are overlapping with each other in terms of their ranges of angles. Two regions that belong to distinct geometrically consistent planes are less likely to have
overlapped angles than those locating within the same plane. With such geometric cues in complement with photometric cues, our method is able to preserve essential geometric regions in the image (Figure 1.4(c)).

Finally, our segmentation method succeeds to extract a variety of compositions from landscape photos, and thus sets stage for many potential composition-based applications ranging from composition-based image retrieval to enhancement of composition design. In this thesis, we employ this technique to build an image retrieval system which provides on-site recommendation of high-quality images with similar compositions to amateur users. Given a query image, which can be a test shot taken by users, we first detect its dominant vanishing point and segment it into slices. By comparing both the location of vanishing point and the segmentation result of the query image with those of all exemplar high-quality images stored in our database, we can select exemplar images with similar compositions to the query image and show users more possible ways of enhancing the aesthetic quality of their photos.

1.5 Modeling Composition in Portrait Photography

Two fundamental questions are often raised when analyzing the composition of a portrait photograph: where are the human subjects in the photo and how do they pose? Traditional composition rules provide us with guidelines to answer the first question. For example, the rule of thirds suggests that putting the human subjects near the 1/3 point of an image is more appealing than at the center. Based on these rules, several methods have been developed to model and assess the positioning of human subjects in a photo.
Nevertheless, the second question remains a challenge. To address this problem, we leverage an important observation in portrait photograph: *experienced portrait photographers often use triangle techniques to create interesting and good-looking poses for human subject*. For example, a widely-used rule for posing is that one should try to avoid 90-degree body angles, because it typically looks unnatural and forced. In addition, triangle techniques are also frequently used to unify multiple human subjects and the surrounding environment, such as chairs and lamps, in a portrait photo.

Despite the popularity of triangle techniques in portrait photography, it is often difficult for amateurs to recognize such triangles, because most triangles do not have explicit edges and sometimes are even constructed by different objects. Moreover, triangles in portraits can be of various sizes, shapes, orientations, and appearances. Hence, our goal is to automatically detect potential triangles from professional photographers’ work in order to help amateurs recognize and learn from the usage of triangle techniques.

Our algorithm can be divided into two steps: First, a line segment detection module is used to extract candidate line segments from an image, which are subsequently filtered using the global contour information in the image. Second, the filtered line segments are fed into a triangle detection module as the candidate sides of triangles. Specifically, a RANSAC algorithm is developed to randomly pick two sides from all the candidates and fit the triangle. Two metrics, *Continuity Ratio* and *Total Ratio*, are defined to evaluate the fitness of these triangles. Only those triangles with high fitness scores will be shown to the users.
1.6 Contributions

Overall, this thesis makes the following major contributions:

- **Composition modeling:** We model the composition of a natural scene image by examining the perspective effects and partitioning the image into photometrically and geometrically consistent regions using our novel hierarchical image segmentation algorithm. We model the composition of a portrait by detecting the lines that can form dominant triangles.

- **Dominant vanishing point detection:** By aggregating the photometric and geometric cues using our segmentation algorithm, we develop an effective method to detect the dominant vanishing point in an arbitrary image.

- **Triangle Detection in Portraits:** We propose a RANSAC algorithm to detect triangles in portraits with a variety of sizes, shapes, orientations, and appearances, helping amateurs to understand and learn from the usage of triangle techniques in professional photographers' work.

- **Composition-sensitive retrieval for on-site feedback:** We develop triangle detection techniques for photographic composition understanding so that photographs with similar or dissimilar composition as the query photo can be retrieved from a collection of photos to assist the photographer.

Admittedly, our technique for natural scenes cannot yet model all potential compositions in such photos, especially when there is a lack of triangles. While in this thesis we focus on the use of triangle techniques in photography, we point out that there are
studies that explore other important aspects of composition, including the semantic features (e.g., buildings, trees, roads) [21, 22, 16, 17]. It would be ideal to integrate all these features in order to gain a deeper understanding of the image composition, but a thorough discussion on this topic is beyond the scope of this thesis.

1.7 Outline of the Thesis

The remainder of the thesis is organized as follows: The related work is discussed in Chapter 2. The methods for modeling composition in natural scenes and portraits are presented in Chapters 3 and 4, respectively. In Chapter 5, experiments and results are described. We conclude and suggest future research in Chapter 6.
Chapter 2

Related Work

In this chapter, we provide an overview of all previous research work related to our thesis. In the first section, we will talk about state-of-the-art techniques currently used to analyze the aesthetic quality of photos based on different types of visual features such as color, light, texture, etc. Then, we will elaborate on how composition of photos can be modeled as well as how this is related to the task of predicting aesthetic quality of photos. In section 2.3, we also give an introduction about 3D modeling and segmentation of natural images. At last, we discuss about portrait photo analysis in section 2.4.

2.1 Photo Aesthetics

One of the earliest work in assessing photo aesthetics [10] characterized dozens of visual features, including colorfulness of photos and purity of these colors, as well as a familiarity measure to evaluate the novelty of the photo based on image retrieval. Ke et al. [26] and Su et al. [56] utilized color histogram to represent the color palette used by a photo. Furthermore, Simplicity [26, 36] and Contrast [12] of color are also taken into account in aesthetic quality assessment. There are other work [36, 35, 50, 40] on combining distinct colors together to generate a more harmonious view.

Aside from color, the use of light can also be a determinant for the aesthetic quality of photographs, especially portraits. Datta et al. [10] used the average pixel
intensity to characterize the use of light in photography. Ke et al. [26] and Luo et al. [36] both pointed out that the brightness of the subject area is significantly different from that of the foreground area, giving rise to a more pleasing look. Dhar et al. [12] focused on outdoor images and introduced three attributes, clear skies, cloudy skies, and sunset skies, to differentiate natural outdoor illumination. Recognizing the importance of lighting condition in human portraits, Luo et al. [35] designed several lighting features such as the ratio of face areas, the average lighting of faces, the ratio of shadow areas, and the face clarity to assess the quality of human portraits.

The low-depth-of-field (low-DOF) technique captures objects within a small range of depths in sharp focus while objects at other depths are blurred. It is usually used to emphasize the subject. Researchers attempted to include the feature of low-DOF in their work to identify high-quality photographs. Some work [10, 24, 50] utilized wavelet-based texture to measure the graininess or smoothness in a photo. Ke et al. [26] computed the spatial distribution of high frequency edges of an image to capture the blur of the background. Dhar et al. [12] employed Daubechies wavelet based features to indicate the blurring amount over the photo. Luo et al. [35] extracted different types of subjects with different approaches such as picking out clear object in a low-DOF photo or detecting humans in a portrait photo. They then measured the clarity of the subject area and treated it as one factor which would influence the aesthetic quality of photographs.
2.2 Composition Modeling

Compared to the above three elements, modeling composition of a photo is more challenging because it requires understanding semantic information. User study indicates composition is the most important feature related to the aesthetic quality of a photo [42, 44]. In photography, composition is the arrangement of visual elements in the scene. Good composition highlights the object of interest and draws people’s attention immediately. Generally, viewers prefer simple and clear compositions. Hence, early work proposed high-level composition features based on locations and orientations of long dominant lines in images [35] to capture the simplicity of the composition. A more intelligent way to describe composition is to model popular composition rules such as the rule of thirds and the rule of golden ratio. The rule of thirds divides the photo evenly with two vertical lines and two horizontal lines, resulting in four intersection points. Studies in photography show that people’s first glances always fall at the four intersections other than the center of the photo. Researchers designed features to model compositions following the rule of thirds [10, 50, 36, 12]. The rule of golden ratio demonstrates that the position of horizon line in scenic photos should be adjusted to satisfy the golden ratio. The golden ratio is always considered as the most beautiful ratio. Bhattacharya et al. [6] enhanced the quality of amateurish photos by adjusting compositions based on golden ratio. Su et al. [56] proposed another novel method to represent different types of photograph compositions. They evenly divided the photo into \( N \times N \) patches where \( N \) equals to \{2, 3, 6\} based on which different patterns of foreground/background area were predefined.
Obrador et al. [42] showed that by using only the composition features, one can achieve image aesthetic classification results that are comparable to the state-of-the-art. Recently, these rules have also been used to predict high-level attributes for image interestingness classification [12] and develop both automatic and interactive cropping and retargeting tools for image enhancement [33, 6]. In addition, Yao et al. [65] proposed a composition-sensitive image retrieval method which classifies images into horizontal, vertical, diagonal, textured, and centered categories, and uses the classification result to retrieve exemplar images that have similar composition and visual characteristics as the query image. However, these features or categories are all defined based on simple 2D rules. Therefore, their ability to capture the variations in image composition induced by the real 3D world is very limited.

2.3 3D Modeling and Segmentation of Natural Images

Various methods have been proposed to extract 3D scene structures from a single image. The GIST descriptor [43] is among the first attempts to characterize the global arrangement of geometric structures using simple image features such as color, texture, and gradients. Following this seminal work, a large number of supervised machine learning methods have been developed to infer approximate 3D structures or depth maps from the image using carefully designed models [21, 22, 16, 53, 39] or grammars [17, 18]. In addition, models tailored for specific scenarios have been studied, such as indoor scenes [31, 20, 19] and urban scenes [4]. However, these work all make strong assumptions on the structure of the scene, hence the types of scene they can handle in practice are
limited. Despite the above efforts, obtaining a good estimation of perspective in an arbitrary image remains an open problem.

Typical vanishing point detection algorithms are based on clustering edges in the image according to their orientations. Kosecká and Zhang [28] proposed an Expectation Maximization (EM) approach to iteratively estimate the vanishing points and update the membership of all edges. Recently, a non-iterative method is developed to simultaneously detect multiple vanishing points [57]. These methods assume that a large number of line segments are available for each cluster. To reduce the uncertainty in the detection results, a unified framework has been proposed to jointly optimize the detected line segments and vanishing points [60]. For images of scenes that lack clear line segments or boundaries, such as the unstructured roads, texture orientation cues of all the pixels are aggregated to detect the vanishing points [48, 27]. However, it is unclear how these methods can be extended to general images.

Image segmentation algorithms commonly operate on low-level image features such as color, edge, texture and the position of patches [54, 13, 32, 2, 38]. However, it was recently shown that given an image, images sharing the same spatial composites can help with the unsupervised segmentation task [51].

2.4 Portrait Photo Analysis

So far, very few studies have been focused on aesthetic analysis of portrait images. Luo et al. [35] studied the critical role of lighting in portrait photography. Varying lighting patterns shift lights and shadows on face, change the area ratios between them,
and generate 3D perception from the 2D photograph. Jin et al. [24] learned four artistic portrait lighting templates, i.e., Rembrandt, Paramount, Loop, and Split, from a database of artistic and daily portrait photographs. These templates are able to capture the composition of low level features. However, we aim at modeling the usage of more holistic composition techniques in portrait photography. Zhang et al. [66] developed a method to automatically recommend suitable positions and poses of people in natural scenes. However, the recommendation is based on matching simple 2D compositional features and manually labeled human body poses.

The pose of human model in portrait photography is strongly related to the whole composition of a photo. In computer vision, human pose estimation is also a challenging problem that lasts many years. Human pose estimation aims at recognizing human poses in videos [67, 52, 14] or still images [59, 15, 63, 64]. Most of these methods focus on single human pose estimation where the scene only contains one human subject. Some of them only consider the upper body of human beings [14, 58, 67]. Recently, Belagiannis et al. [5] proposed an approach to estimate jointly poses for multiple human beings in a multi-camera environment. However, in portrait photography, a common scenario is that we need to take pictures for two or more people such as couple or family with only one single camera. Furthermore, professional photographers not only embed triangles in human poses but also design the composition in a way everything in the scene, e.g., apparels, shadow, chairs, walls, etc., can play a role in constructing triangles.
Chapter 3

Modeling Composition in Natural Scene Photography

As mentioned in chapter 1, the pie model enables us to describe a variety of different compositions in natural scene photography. In order to detect the pie, we propose a hierarchical image segmentation algorithm to partition images into pie slices. In this chapter, we first briefly introduce the framework of hierarchical image segmentation. Then, in Section 3.2, we discuss into details about the photometric and geometric distance measures used in the framework, assuming the location of the dominant vanishing point is known. In Section 3.4, we further propose a novel approach to enhancing vanishing point detection by aggregating photometric and geometric cues within regions. Finally, we apply this pie detection technique to retrieve and recommend exemplar photos with similar compositions to users.

3.1 Hierarchical Image Segmentation

Hierarchical image segmentation is a bottom-up method developed based on agglomerative clustering. In the beginning, each pixel in an image is considered as a single region/cluster. Out of all these regions/clusters, a pair of adjacent regions/clusters with minimum distance is selected and merged to generate a bigger region/cluster iteratively, until the stop criterion is satisfied. Here, the distance of two regions is defined to measure the likelihood that two regions belong to the same geometrically consistent plane. The
selection of stop criterion is relatively flexible. For instance, the process can stop when a fixed number of regions is achieved or the minimum distance exceeds a certain threshold. Each iteration of hierarchical image segmentation can be formalized into three steps:

1. Select a pair of regions with minimum distance: $d^* = \min_{i,j} d_{ij}$.

2. Let $R_1, R_2 \in \mathcal{R}$ be the regions with distance $d^*$. Set $\mathcal{R} \leftarrow \mathcal{R} \setminus \{R_1, R_2\} \cup \{R_1 \cup R_2\}$ and update distances between other regions and the newly generated region accordingly.

3. Return all the regions if the stop criterion is satisfied. Otherwise, go back to step 1.

To improve the efficiency of clustering, we do not use each pixel as an initial region. Instead, we apply over-segmentation techniques to generate small regions and start agglomerative clustering from these regions. We select Arbelaez’s segmentation algorithm [2]. Compared to other existing methods [13, 38] utilizing local cues to measure visual dissimilarity between regions, Arbelaez et al. [2] propose to integrate global photometric information into contour detection via spectral clustering. It helps identifying contours with weak local cues. Since the detected contours are all closed, segmentation result can be easily inferred from contours. The result of Arbelaez’s method is an Ultrametric Contour Map (UCM) [1], shown in Figure 3.1(b). Each pixel on contour is assigned with real number from 0 to 1, called a confidence value, indicating the dissimilarity between two regions separated by the contour. Darker pixels in Figure 3.1(b) imply higher confidence value, i.e., the photometric dissimilarity between two adjacent regions. We employ over-segmented UCM result as initial regions in our framework (i.e.,
by thresholding the UCM at a small scale 0.05). We also use the confidence value of contour as photometric distance between two regions during merging iterations. However, photometric cues in photos sometimes are too weak to separate geometrically consistent regions (Figure 3.1). As a result, we propose a geometric distance measure and linearly combine both distances to measure the likelihood that two adjacent regions lie in the same geometric plane.

![Image](image.png)

(a) The original image. (b) The ultrametric contour map (UCM) generated by Arbelaez et al. [2]. (c) The segmentation result obtained by thresholding the UCM at a fixed scale.

3.2 Geometric Distance Measure

The over-segmented ultrametric contour map presents a number of initial regions. Any two regions share a segment of contour. The confidence level of the contour indicates photometric dissimilarity between the two adjacent regions while the orientation of the contour to some degree reveals the relative geometric relationship between them. Consider the pie model in Figure 1.3, several parallel lines converge to one dominant
vanishing point. These parallel lines are the boundaries between two geometrically consistent planes (e.g., the horizon between sky and roads). Therefore, if two adjacent regions share a boundary passing through the dominant vanishing point, they are likely to be lying on two different geometric planes. However, in real world scenarios, some geometrically consistent planes do not share perfectly straight boundaries passing through the dominant vanishing point (e.g. the indented boundary between wall and ceiling in Figure 3.2). If we focus our attention on the three adjacent regions $R_1$, $R_2$, and $R_3$ in Figure 3.2, we notice that $R_1$ and $R_3$ belong to the vertical wall and $R_2$ belongs to the ceiling. However, the boundaries between the pair $(R_1, R_2)$ and the pair $(R_1, R_3)$ both lie on the same (vertical) line. As a result, it is impossible to differentiate these two pairs based on only the orientation of these boundaries.

In order to solve this problem, we first transform the image into a polar coordinate system with its origin at the dominant vanishing point $P$ as shown in Figure 3.2(a). Given any pixel $X$ within the image, its polar coordinates can be represented as $(r_X, \theta_X)$, $\theta_X \in [0, 2\pi)$ where $r_X$ is the distance from pixel $X$ to the origin and $\theta_X$ is the angle between line $OX$ and the positive horizontal axis. We compute the histogram of the angle values $\theta_X$ for all pixels $X \in R_i$, as shown in Figure 3.2(c). The histogram describes how region $R_i$ distributes across angles from 0 to $2\pi$. Apparently, two adjacent regions falling in separate geometrically consistent planes have their pixels distributed over different ranges of angles. On the contrary, two regions with highly-overlapped angle coverage are more likely to lie in the same geometric plane.
Based on this observation, we define the geometric distance between any two regions $R_i$ and $R_j$ as follows:

$$d_{ij} = 1 - \max \left( \frac{\sum_{\theta} \min(c_i(\theta), c_j(\theta))}{|R_i|}, \frac{\sum_{\theta} \min(c_i(\theta), c_j(\theta))}{|R_j|} \right)$$

(3.1)

where $|R_i|$ and $|R_j|$ are the total numbers of pixels in regions $R_i$ and $R_j$, respectively. $c_i(\theta)$ and $c_j(\theta)$ are counts of the $\theta$-th bin in histograms for region $R_i$ and $R_j$, respectively. Therefore, $\sum_{\theta} \min(c_i(\theta), c_j(\theta))$ computes the overlapped area of the two histograms. Its value divided by the size of $R_i$ and $R_j$ indicates how the two histograms, i.e., distributions of polar angles, overlap with each other overall. If the histograms of $R_i$ and $R_j$ have very small overlapped area, they are unlikely to be in the same geometric plane. Thus, the geometric distance $d_{ij}$ is close to 1. Conversely, highly-overlapped histograms imply two adjacent regions lying on the same geometric plane and result in a small $d_{ij}$. For example, as illustrated in Figure 3.2(c), $R_1$’s histogram covers the entire histogram of $R_3$ and hence we have $d_{13} = 0$. Meanwhile, $R_1$ and $R_2$ have very small overlap in their histograms. As a result, their geometric distance is large: $d_{12} = 0.95$.

In Figure 3.2(d), we show the boundary map weighted by the geometric distance between adjacent regions. As expected, the boundaries between geometrically consistent planes (e.g., the boundary separating wall and ceiling) are weighted heavier than those within geometric planes. Compared to Figure 3.1(c) where photometric cues are used to assign weights on contours, the geometric cues are good at separating planes with little noticeable photometric dissimilarity.
3.3 Detecting Pie by Integrating Photometric and Geometric Cues

In Chapter 3, we have introduced the two types of distance measures used in our hierarchical image segmentation: photometric distance and geometric distance. By linearly integrating them, we are able to detect both photometrically and geometrically consistent planes in an image, i.e., the triangular slices in the pie model.

Formally, given two adjacent regions $R_i$ and $R_j$, the distance between them is defined to be a linear combination $d_{ij} = \lambda d_{ij}^p + (1 - \lambda)d_{ij}^g$, where $d_{ij}^p$ is photometric distance, assigned as the confidence level of contour separating the two regions and recorded in ultrametric contour map [2]. $d_{ij}^g$ is geometric distance defined in Equation 3.1, indicating the geometric relationship between two regions with respect to the dominant vanishing point. By adjusting the weight $\lambda$, we can determine which type of cue is emphasized more than the other.
Fig. 3.3 Image segmentation results by integrating both photometric and geometric cues. Different weighting parameter $\lambda$ have been used.

Figure 3.3 presents several examples of detected pie with different $\lambda$ and a fixed number of regions $K$. $\lambda = 1$ means the hierarchical segmentation framework only includes geometric cues. The algorithm successfully extract most geometric structures from the scene in spite of similar photometric cues across different regions. However, it made mistakes on some small parts that can be easily separated with photometric cues (e.g., the upper part of the left bridge and the sky). When $\lambda$ equals to 0, only photometric cues are taken into account. Some regions with similar photometric cues are mistakenly merged into one even though they do not belong to the same geometric plane. To take both types of cues into account, the best choice of $\lambda$ is between 0.4 and 0.6. For the following experiment, we empirically set $\lambda$ to be 0.4. Figure 3.4 provides more segmentation results.
Fig. 3.4 Additional segmentation results. For each original image, we show the results with $\lambda =1, 0.8, 0.6, 0.4, 0.2$ and 0, in this order respectively.
3.4 Enhancing Vanishing Point Detection

In Section 3.2, we have shown that the location of dominant vanishing point helps identifying geometrically consistent planes in the pie model. However, detecting the dominant vanishing point itself is a challenging problem in computer vision. Most present methods extract a number of line segments from the image that converge into a single point and label that point as a potential vanishing point. Such methods work well in man-made environment where enough line segments can be extracted from buildings or grounds. With regard to natural scene photography, there does not exist perfectly straight line segments. For instance, the photo in Figure 3.5(a) does not contain strong straight line cues (Figure 3.5(b)). It is still easy for human beings to recognize the dominant vanishing point in the photo. However, according to Figure 3.5(b), existing vanishing point detection mechanism heavily depends on line cues and thus fails to identify the correct vanishing point.

It is worth noting that although no straight converging lines are present in the image, geometric cues will still aggregate along the vanishing directions which converge towards the correct dominant vanishing point (Figure 3.5(c)). Specifically, given a hypothesis for the dominant vanishing point, we can always calculate a boundary map with respect to the vanishing point. The closer the hypothesis vanishing point locates from the correct one, the more geometric cues will aggregate along the boundaries. That is to say, the resulting boundary map contains more strong boundaries with larger distances between adjacent regions. Our algorithm can be summarized as follows:

*In this thesis, we assume that the dominant vanishing point lies in the image frame because, as we noted before, only the vanishing point which lies within or near the frame conveys a strong
Fig. 3.5 Enhancing vanishing point detection. (a) Original image. (b) Line segment detected. (c) and (d) The weighted boundary map for two different hypotheses of the dominant vanishing point location. (e) The consensus scores for all vertices on the grid.

1. Uniformly divide the image by an $m \times n$ grid. Each grid holds a hypothesis of dominant vanishing point $P_{kl}$ in the middle where $0 \leq k \leq m - 1$, $0 \leq l \leq n - 1$.

2. Given each hypothesis, we compute the geometric distance $d_{ij}^B$ for any pair of regions $R_i$ and $R_j$ in an over-segmentation of the image. Then, a consensus score for $P_{kl}$ is defined as: $f(P_{kl}) = \sum_{ij} d_{ij}^P \times d_{ij}^B$.

3. Select the point with the highest score as the detected dominant vanishing point: $P^* = \arg \max f(P_{kl})$.

Figure 3.5(c) and 3.5(d) provide boundary maps for both the correct and wrong hypothesis of dominant vanishing points. Apparently, correct hypothesis gains stronger geometric cues and thus clearer boundaries than wrong hypothesis. Figure 3.5(e) exhibits sense of 3D space to the viewer. Nevertheless, our method can be easily extended to detect vanishing points outside the frame using a larger mesh grid.
consensus scores for all hypothesis vanishing points within the grid mesh. The brightest point indicates a highest consensus score.

Here, the size of the grid may be chosen based on the desired precision for the location of the vanishing point. In practice, our algorithm can find the optimal location in about one minute on a 50 × 33 grid on a single CPU. We also note that the time may be further reduced using a coarse-to-fine procedure.
In this chapter, we describe the algorithm for handling portrait photos. The idea of modeling compositions for portrait photographs is to find out the embedded triangles and how they influence the way we comprehend the entire scene. Our algorithm consists of three steps: line segment detection, triangle construction, and triangle evaluation. We first introduce the line segment detection algorithm, and then discuss how triangles can be constructed from these line segments. Finally, two scores, i.e., Continuity Ratio and Total Ratio are defined to evaluate the fitting quality of a triangle, based on which we make a decision of keeping the detected triangle or not. We also develop a potential application of our algorithm: a triangle retrieval system which detects embedded composition-relative triangles in professional portrait photographs to help amateur users gain deeper understanding about composition design skills.

4.1 Line Segment Detection

By examining high-quality portraits designed with the triangle technique (Figure 4.1), we observe that triangles present in portrait photographs are often composed with parts of contours or edges, such as the contours of arms, body parts, wearing apparels, as well as edges formed by multiple human subjects. In general, a triangle is geometrically defined as a polygon with three corners and three sides, where the three...
Fig. 4.1 Examples of implied triangles in portrait photography.

sides are all straight line segments. However, contours of natural objects like humans or hats are often slightly curved. Therefore, to detect potential triangles in real images, our method needs to be able to identify such slightly curved line segments, in addition to straight edges, in an image.

To this end, we employ the Line Segment Detector (LSD) [62] to convert the gradient map of an image to a set of line segments. The line segment detector works as follows. It first calculates a level-line angle at each pixel to produce a level-line field. The level line is a straight line perpendicular to the gradient at each pixel. Then, the image is partitioned into line-support regions by grouping connected pixels that share the same angle up to a certain tolerance. Each line-support region is treated as a candidate line segment. Next, a hypothesis testing framework is developed to test each line segment candidate. The framework approximates each line-support region with a rectangle and compare the number of “aligned points” in each rectangle in the original image with the expected number of aligned points in a random image. A line segment is detected if the actual number of aligned points in a rectangle is significantly larger than the expected number.
Fig. 4.2 LSD results with different density.

A nice property of LSD is that, by approximating the line-support region using a rectangle of certain length, it is able to detect near-straight curves in an image. Further, it is straightforward that a larger rectangle with more unaligned points will be needed to cover a more curved line segment. Therefore, by setting a threshold on the density of a rectangle, which is defined as the proportion of aligned points in the rectangle, we can control the degree up to which a curved line segment is considered. Figure 4.2 shows results of the line segment detector with different density values. The density is empirically set to 0.2 in our experiments.

While the line segment detector aims at extracting all potential line segments from an image using local image gradient cues, we notice that some line segments are more globally distinguishable and thus more visually attractive to viewers than others. To further identify such line segments, we combine the line segment detector with the Ultra-metric Contour Map obtained by the state-of-the-art image segmentation algorithm [3]. As we discussed in Section 3.1, the contour map has the same size as the original image. Each pixel on the map holds a confidence level between 0 and 1, indicating the possibility
of it being on a boundary. Therefore, given a line segment, we identify all the pixels falling in its support region and consider the maximum confidence level of all pixels as the confidence level of the line segment. Finally, line segments with confidence levels under a certain threshold are discarded. We choose the threshold based on the maximum confidence level present in an image. Specifically, assume the maximum confidence level of all the line segments in an image is $C$, where $C \in [0, 1]$, then we set the threshold as $(1 - \alpha)C$ and accept line segments whose confidence levels are within the range of $[(1 - \alpha)C, C]$. The parameter $\alpha$ controls the number of accepted line segments. Smaller $\alpha$ filters out more line segments from an image, as shown in Figure 4.3.

4.2 Detecting Triangles

The line segment detector described above gives us a set of candidate triangle sides. Randomly selecting three non-parallel sides from the set generates a triangle.
Therefore, the problem of detecting a triangle can be converted into finding three non-parallel sides. However, although a triangle consists of three sides, we observe that a triangle can be uniquely determined as long as two sides are found, because the third side can be obtained by connecting the end points of the other two sides. Moreover, the presence of the third side is not as important in the practical usage of triangle technique because viewers can easily “complete” the geometric shapes themselves. As a result, our problem is reduced to fitting a triangle using two non-parallel line segments selected from the candidate set.

Two major challenges still exist in fitting triangles using the extracted line segments: 1) there are a large number of outlying line segments; and 2) the sides of a triangle are imperfect in real images. For example, as shown in Figure 4.4(a), some objects in a photograph are irrelevant to the use of triangle technique, even though they have high-contrast contours (e.g., the bird pattern on the woman’s shirt). The line segments produced by such objects are all outliers. In addition, multiple triangles often exist in one image (Figure 4.4(b)). Thus, line segments from different triangles should
also be considered as outliers with regards to each other. In Figures 4.4(c) and 4.4(d), we further show some examples of imperfect triangle sides. As one can see, these sides may be broken into several parts because of occlusion or artificial effect introduced by the line segment detector. For instance, the girl in Figure 4.4(c) has her right arm occluded by her left arm. As a result, the contour of her right arm is broken into two line segments. In Figure 4.4(d), the contour of the back of the human subject is too curved to be approximated by a single straight line. Therefore, the line segment detector approximates it using two straight line segments with slightly different orientations.

In order to tackle these two challenges, we employ a modified RANSAC (RANdom SAmple Consensus) algorithm in favor of its insensitivity to outliers. RANSAC is an iterative method that robustly fits a set of observed data points (including outliers) to a pre-defined model. Our modified RANSAC algorithm includes three steps:

1. Two non-parallel line segments are randomly selected from the candidate set and are extended to generate two lines on which the two triangle sides lie.

2. All the candidate line segments within neighborhoods of the two lines are projected onto the correspond lines, resulting in a number of projected pixels. Triangle sides are then constructed from all the projected pixels.

3. Once two triangle sides are constructed, two metrics Continuity Ratio and Total Ratio are calculated to measure the fitness and significance of the triangle, respectively. Triangles with high scores are accepted.

Below we describe each step in details.
4.2.1 Identifying Sides From Line Segments

By extending the two randomly selected line segments to two lines, we determine the shared end point of the two sides, which is the intersection of the lines. Moreover, two intersecting lines generate four different angles with four different opening directions: upwards, downwards, leftwards, and rightwards. Each angle corresponds to a category of possible triangles that contain this angle and two sides of varied lengths. Given one of the four angles, once the lengths of its two sides are determined, a unique triangle can be constructed.

4.2.2 Fitting All Segments on Sides

In this step, we first mark all line segments within neighborhood of the two straight lines as inliers and those falling outside neighborhood as outliers. The neighborhood region of a straight line \( l \ (ax + by + c = 0) \) is defined to be \( N(l) = \{(x,y)|(x,y) \in \mathbb{R}^2 \text{ and } \frac{|ax+by+c|}{\sqrt{a^2+b^2}} \leq d_{nb}\} \), i.e., the group of pixels whose distances to the straight line are smaller than a certain threshold \( d_{nb} \). Then, all the inlier line segments with respect to
line $l$ can be calculated as $I(l) = S \cap N(l)$ where $S$ is the set of all candidate line segments. Note that, if a line segment is cut into two parts by the neighborhood boundary, the part of line segment falling within the neighborhood is included as an inlier, whereas the other part is considered as an outlier. In Figure 4.5(c), we illustrate how the neighborhood of a straight line is utilized to partition all the line segments into inliers and outliers. The brown line segment is selected from the candidate line segment set and extended to a straight line $l$. Lines $l_1$ and $l_2$ designate boundaries of the neighborhood region. Both of them have a distance $d_{nb}$ from $l$. The red line segments and black line segments represent the inliers and outliers, respectively.

Next, all the pixels on the inlier line segments are projected onto the straight line. A pixel on the straight line is called a projected pixel if there is at least one pixel on any inlier line segment which is projected to this pixel. We denote the set of all projected pixels as $P(l) = \{(x', y') \in l \mid \exists (x, y) \in I(l) \text{ and } (x - x', y - y') \perp l\}$.

Figure 4.5 illustrates the fitting process. Figure 4.5(b) shows all the line segments extracted from the image shown in Figure 4.5(a). The two line segments colored in blue are randomly selected from the candidate set. Subsequently, two straight lines, $l$ and $\tilde{l}$, are generated by expanding the two line segments and their neighborhood regions are identified. Finally, inliers within their neighborhoods are projected onto the straight lines, as shown in Figure 4.5(d).

Here, we note that the projected pixels typically scatter along the entire straight line. However, a triangle is formed by two half lines determined by the intersection point. Therefore, when evaluating the fitting qualify of a triangle, we only consider the
subsets of projected pixels which are on the two half lines, denoted as \( P(l^h) \) and \( P(\bar{l}^h) \), as opposed to the entire sets of projected pixels \( P(l) \) and \( P(\bar{l}) \).

### 4.2.3 Evaluating the Fitted Triangle

In order to evaluate the quality of a fitted triangle, we first introduce a Continuity Ratio score to evaluate the quality of a fitted side. Here we use one half line \( l^h \) as an example. The way to calculate continuity ratio for the other half line \( \bar{l}^h \) is exactly the same. Given any point \( X(x, y) \) lying on the half line \( l^h \), we can construct a potential side \( OX \) connecting the intersection point \( O(x_o, y_o) \) and the given point. The Continuity Ratio for this potential side is defined as follows:

\[
C(OX) = \frac{|P(l^h) \cap OX|}{|OX|}
\]

The next step is to find the best fitted side on \( l^h \):

\[
\max_{X \in P(l^h)} C(OX)
\]

Similarly, we can find the best fitted side \( O\bar{X} \) for the other half line \( \bar{l}^h \). The continuity ratio for the entire triangle constructed by the two sides is \( C(OX) \times C(O\bar{X}) \). In addition to the continuity ratio, we define another Total Ratio score which is calculated as the area of the triangle divided by the area of the entire picture. Specifically, the Continuity Ratio is in fact the product of the ratio of projected pixels along each side. This score describes how well the extracted line segments fit the given side. Meanwhile, the Total Ratio represents the significance of a fitted triangle in terms of sizes. As a bigger triangle
can be more easily recognized and has more impact on composition of the entire image, we only keep the triangles whose Continuity Ratio and Total Ratio scores are both above certain thresholds.

4.3 Portrait Retrieval using Triangles

Our triangle detection technique can potentially help amateur photographers to design poses when taking pictures, because it is usually difficult for them to naturally embed triangles in compositions to form high-quality portraits. Magazine editors can also benefit from the triangle detection technique. When they are selecting portraits to fill a specific page layout, the shapes and orientations of embedded triangles within these photos can be critical to the overall composition. Therefore, we develop a portrait retrieval system that can take a triangle sketch as a query to enable users to find images containing specific triangular configuration.

First, users need to provide a sketch indicating the shape and the orientation of the triangular configuration that they are looking for. In portrait photos, the third side of a triangle is often missing in most photos. Thus, the sketch query provided by users is basically an angle with two sides. Given the sketch angle, the orientations of the two sides as well as the direction of the angle can be calculated and used for further searching. Specifically, the orientation of a side is defined to be the angle between its extended straight line and the positive $x$-axis. An angle has four possible directions: upward, downward, leftward, or rightward. The two properties narrow down the searching space to a specific type of triangles.
Recall that our triangle detection algorithm contains two steps: line segment detection and fitting triangles. All line segments detected from the first step are taken as candidate triangle sides during the fitting stage. However, not all line segments are relevant to the specific type of triangles that we are looking for. In the retrieval system, we construct two candidate sets containing line segments with similar orientations as the two sides of the sketch angle. Two sides can then be randomly selected from the two candidate sets and the combination of them generates four angles with four different directions (Figure 4.5). Only the angle that has the same direction as the sketch angle will be taken into consideration during the fitting process. For instance, if the sketch angle is facing upwards, the fitting process will only fit triangles containing the upward angle as in Figure 4.5(d). Such a sketch-based triangle retrieval system not only assists users in searching for a specific type of triangles but also reduces searching time a lot.
Chapter 5

Experiment

In this chapter, we construct two datasets and evaluate our algorithms for both categories of photos. For natural scene photos, we evaluate the performance of our pie detection algorithm, vanishing point detection algorithm, and composition-sensitive natural scene image retrieval system. For portrait photos, we also evaluate our triangle detection algorithm and further discuss the performance of our triangle-based portrait retrieval system.

5.1 Evaluation of Pie Detection

To evaluate the performance of our pie detection algorithm, we manually labeled 200 images downloaded from flickr.com using the labeling tool provided in Arbelaez’s work [2]. While manually partitioning the image into different segments, our goal is to separate regions with different photometric cues that are lying on different geometric planes. Figure 5.1 shows eight examples of manually labeled segmentations.

We compare the performance of our method with the state-of-the-art image segmentation method, gPb-owt-ucm [2]. In our experiment, we assume the location of the dominant vanishing point is known. We need to point out that our goal here is not to compete with Arbelaez’s algorithm but to demonstrate the advantage of utilizing vanishing point in addition to photometric cues in detecting the pie model.
To quantitatively evaluate the performance of our algorithm, we adopt three popular metrics: Rand Index (RI), Variation of Information (VOI), and Segmentation Covering (SC). The three metrics are designed to compare the test segmentation result $S$ with the ground truth segmentation $G$ of an image. The RI metric calculates the probability of any arbitrary pair of pixels that have the same label in $S$ and $G$ or have different labels in both segmentations. The range of the RI metric is $[0, 1]$, with higher values indicating greater similarity between two partitions. The VOI metric measures the distance between two segmentations in terms of their average conditional entropy. Smaller VOI values imply a closer test segmentation result to the ground truth. The SC metric investigates the degree to which regions in $S$ cover regions in the ground truth segmentation $G$. As Rand Index, SC also ranges in $[0, 1]$, with higher values indicating more accurate test segmentation results. We refer readers to Arbelaez’s work [2] for more details about these metrics.

Figure 5.1 also shows the benchmark results of both methods. As one can see, our method significantly outperforms $gPb$-owt-ucm on all metrics. This is consistent with the example segmentation results we show in Figures 3.3 and 3.4, which clearly suggest that our method is advantageous in segmenting the geometric structures in the scene.

### 5.2 Vanishing Point Detection

In this section, we compare our vanishing point detection method with two state-of-the-art methods proposed by Tardif [57] and Tretiak et al. [60]. As discussed earlier, both methods heavily rely on the presence of line segments in images. Tardif [57] first extracts line segments from images and then develops a non-iterative scheme to group
Fig. 5.1 Segmentation benchmarks. $K$ is the number of regions. Top rows show some example test images with the manually labeled segmentation maps.

to evaluate the performance of our method, we randomly picked 400 images from our database whose dominant vanishing points lie within the image frame. To make the comparison fair, for both work [57, 60] we only keep the vanishing point with the largest support set among all hypotheses that also lie within the image frame. We consider a detection to be successful if the distance between the detected vanishing point and
the manually labeled ground truth is smaller than 20 pixels. The success rates for Tardif [57], Tretiak [60], and our method are 58.0%, 61.8%, and 70.0%, respectively, which suggest that our method is more accurate for detecting the dominant vanishing point in arbitrary images. We also note that while the joint optimization scheme [60] is effective in recovering weak line segments and vanishing points for urban scenes, its improvement over Tardif’s algorithm [57] is quite small in our case.

5.3 Composition-Sensitive Image Retrieval

As discussed above, we are able to extract pie composition model from a photo with a hierarchical image segmentation method using both photometric and geometric cues. Correctly extracting compositions from photographs motivates a number of potential composition-based applications. For instance, it enables us to evaluate the aesthetic quality of a composition, edit a composition by manipulating the extracted pie, and retrieve images with similar compositions. In this thesis, we build an image retrieval system that aims at searching for images with similar compositions to a query image from a large gallery of high-quality images. The goal is to recommend exemplar photos with specific compositions to amateur users. Yao et al. [65] pioneered this direction, but the types of composition studied there are limited to a few categories which are pre-defined based on simple 2D rules.

Compared to Yao’s work, our system does not divide compositions into limited number of categories. Instead, it is able to model any specific composition by identifying two important compositional elements: the location of the dominant vanishing point and how each segment relates to the point. Professional photographers are skillful at
adjusting these two elements to generate a variety of high-quality compositions. Therefore, in order to compare two compositions, we need to compare both the locations of dominant vanishing points and how the scenes are segmented with respect to the point. Being aware of this, we further develop a similarity measure to compare two compositions based on their geometric segmentation maps. Given two images $I_i$ and $I_j$, we denote the locations of their dominant vanishing points as $P_i$ and $P_j$. Also, our method can generate segmentation maps $S_i$ and $S_j$ from the two images. The similarity measure is defined as follows:

\[
D(I_i, I_j) = F(S_i, S_j) + \alpha \|P_i - P_j\|,
\]  

(5.1)

where $\|P_i - P_j\|$ calculates the euclidean distance between two points and $F(S_i, S_j)$ is a metric to compare two segmentation maps. We adopt the Rand index [47] for its effectiveness. In addition, $\alpha$ controls the relative impact of the two terms in Eq. (5.1). We empirically set $\alpha = 0.5$.

In Figure 5.2, we show various examples of query images as well as retrieved results. For this experiment, we collect 3,728 images from flickr.com by querying the keyword “vanishing point”. Each image in this database is scaled to size 500 × 330 or 330 × 500. We manually label the dominant vanishing point and then apply the hierarchical image segmentation algorithm (with the proposed dissimilarity measure and the stopping criteria $\delta = 0.55$) to obtain a segmentation for each image. The results clearly show that the proposed measure is not only able to find images with similar dominant vanishing point locations, but also effectively captures how each region in

*Here, we assume the two images have the same size. In practice, we scale all the images to the same size for easy comparison.
Fig. 5.2 Composition-sensitive image retrieval results. Each row shows a query image (leftmost) and the top-6 or top-8 ranked images retrieved.
the image is related to the vanishing point. For example, the images in the 4th, 6th, and 7th rows of Figure 5.2 all have similar vanishing point locations (around the image center) but very different scene structure compositions, and thus convey very different impressions to the viewer.

5.4 Triangle Detection in Portrait Images

To the best of our knowledge, we are the first to detect the use of triangle techniques in portrait photographs. In order to evaluate the performance of our triangle detection system, we build up a dataset by collecting 4451 professional photographs from Flickr. We use the keyword “studio portrait” for search because studio portraits are often taken by professional photographers who can use the triangle techniques skillfully and embed various triangles in their works. Four examples of portrait photographs which use triangle techniques are shown in Figure 4.1.

Figure 5.3 presents detected triangles with different continuity ratios. We empirically set the thresholds as 0.7 for Continuity Ratio and 0.05 for Total Ratio. Specifically, given a fitted line segment, if more than 30% of its points are missing, we will remove it from the final detected results. As a result, the continuity ratios for a whole triangle (including two sides) is above $0.7 \times 0.7 = 0.49$. The threshold 0.05 for total ratios means that triangles with areas smaller than 5% of the entire area of the photo are not considered to be significantly related to the whole composition. We will also eliminate such triangles from final detected results. As one can see from Figure 5.3, triangles with high continuity ratios are more easily recognized than those with low continuity ratios. However, they do not necessarily outperform those of low continuity ratios in
conveying useful compositional information about the photo. For instance, Figure 5.3(g) has a much lower continuity ratio than Figure 5.3(j) but it conveys a more interesting compositional skill. From the detected triangle, we can notice that the model slightly tilts her head to align with her left arm which constructs a beautiful triangle with her hair. It is very common that multiple different triangles are embedded in one image, and some of them can be easily overlooked by amateurs. Our goal here is to identify all potential triangles from images, which enables amateur photographers to better learn useful compositional techniques.

More detected results can be found in Figure 5.4. It demonstrates that our triangle detection system can clearly identify presence of triangles in portrait photographs despite the existence of noises such as human hair, props, and shadows/patterns/folds on shirts, etc. Moreover, triangles involving multiple human subjects can be accurately detected as well (e.g., the fourth picture in the first row and the third picture in the second row). More interestingly, our system is able to locate triangles that may not be easy to identify by human eyes. For instance, considering the first photo in the last row, it is quite easy to find the triangle containing the girl’s two arms. However, we often overlook another triangle which is constructed with one arm of the girl and the edge of her lower jaw. Another example is the fifth picture in the last row. The girl puts her arm on her dress in a deliberately designed pose so that it extends the boundary of her dress and forms a big triangle together with her long hair. Both examples indicate that professional photographers usually design delicate poses for the subjects in order to achieve good photo composition. However, choosing a good pose requires much experience and
Fig. 5.3 Detected triangles with different Continuity Ratios
Fig. 5.4 Examples of detected triangles in portrait photographs.

artistic inspiration. The triangles detected by our system can help amateurs gain deeper understanding and inspirations from high-quality photography works.

5.5 Portrait Retrieval using Triangles

Further, to evaluate the performance of the triangle retrieval system, we select 20 groups of representative queries which cover a wide range of angles in terms of magnitude and orientation. We only consider angles in the range of \([45^\circ, 135^\circ]\) because both too small and too big angles are not recognized as triangles. Specifically, each of the 20 groups of queries takes a distinct combination of orientations for two straight lines. Twenty line combinations \(\langle l_1, l_2 \rangle\) are selected in our experiment such that the angle between \(l_1\) and \(l_2\) falls in the closed range of \([45^\circ, 135^\circ]\) and the angle between \(l_1/l_2\) and positive \(x\)-axis falls
in \( \{0^\circ, 22.5^\circ, 45^\circ, 67.5^\circ, 90^\circ, 112.5^\circ, 135^\circ\} \). Moreover, one combination of two straight lines generates four possible angles which differ in terms of their opening directions. Therefore, four different opening directions are included within each group, namely, upward, downward, leftward, and rightward. We use the 80 queries to retrieve triangles from 4451 photos where each photo may contain many distinct triangles. For each query, we rank the results simply based on their continuity ratios. Higher continuity ratios represent higher quality of fitting and thus imply more accurate retrieved results. That is to say, triangles with higher continuity ratios are less likely to be noise, i.e., the two sides are indeed present in the photos. However, more accurate triangles are not necessarily more “useful” in terms of conveying valuable information about composition. Our goal is to identify as many accurate triangles as possible with the assumption that some of them may provide useful guidance. More work can be done in the future to evaluate and predict the usefulness of a specific triangle. To gain a brief sense about the advantage of our system, we take the top 20 results for each query and evaluate these results by manually labeling whether each result is “useful”. More specifically, we check whether a retrieved result conveys information about how professional photographers design the composition of a photo because our system aims at helping amateur users discover such information. It turns out 73.21% of the retrieved results are delivering composition-related information.

Figure 5.5(a) shows how the performance of our portrait retrieval system changes as the number of retrieved results increases from 1 to 20. The performance is calculated as the percentage of results that are labeled as “useful” in terms of providing composition-related information. We can notice that the performance of our system is quite stable
Fig. 5.5 The performance of portrait retrieval system.

at 73%. In addition, if we look at the histogram of performance for 80 different queries in Figure 5.5(b), it demonstrates that most queries have 70% - 90% of retrieved results that can deliver useful information to help amateur users.

Furthermore, to demonstrate how our system works in more details, we provide examples of both “useful” and “useless” retrieved triangles in Figure 5.6. The first column contains twelve queries among which eight return “useful” triangles and four return “useless” triangles. Whether a retrieved triangle is useful depends on whether it provides information about the composition skills of professional photographers. For each query, we only take into account the orientations of two sides and ignore information such as its length and width. It is reasonable because amateur users often do not have a clear idea about how to embed triangles in photos, let alone provide precise description about the triangles they are looking for. This approach enables amateur users to find more diverse triangles and thus can provide users effective suggestions on mastering the triangle composition technique. From the results of our system, we can learn that
professional photographers are good at using all kinds of objects, such as arm, leg, shoulder, chair, wall, ground, hair, apparel, or even shadow, to construct triangles. It is worth noticing that slight adjustment of a pose can form beautiful triangles which make the entire composition aesthetically appealing. For instance, the girl in row 3, column 3 slightly turns her head towards left to perfectly align with her shoulder. Amateur users will not pay attention to such details. With our system, they are able to recognize the intricacy of a designed pose. It is due to the same reason that the girl in row 1, column 4 lifts up her head a little bit. Our system also retrieves “useless” triangles for some queries. These triangles seem not to be interesting at all but they also convey useful information. As mentioned in Section 1.5, professional photographers sometimes will avoid 90-degree body angles because it typically looks unnatural and forced. It has been verified by our system that combinations of 0° and 90° sides do not construct aesthetic triangles. It can be seen that although our system can correctly detect 90-degree angles from photos, most of these angles are just noise and do not hold high aesthetic quality.
Fig. 5.6 Useful and useless examples of retrieved portraits. Each row contains results from the same query.
Chapter 6

Conclusions and Future Work

This thesis proposes a system to detect the usage of triangle techniques in both natural scene and portrait photography and use it to improve the composition of photographs. For photographs of natural scene with strong perspective effects, we present a new method to model visual composition through analyzing the perspective effects and segmenting the image based on photometric and geometric cues. The method can effectively detect the dominant vanishing point from an arbitrary scene. For portrait photographs, we extract a set of candidate line segments from a photo and then successfully fit triangles to these segments despite a large proportion of outliers. The fitted results accurately identify the presence of triangles in photographs. Among a variety of potential applications, we have illustrated how our techniques can provide on-site feedback to photographers.

Our work opens up several future directions. First, we may extend our geometric image segmentation and vanishing point detection algorithms to handle images with two or more dominant vanishing points, which can often be found in man-made environments. In such case, our goal is to detect all the dominant vanishing points in an image and to group the regions according to the corresponding vanishing points. Another challenge in composition recognition for real-world photos is the presence of large foreground objects. They typically correspond to regions which do not associate with any vanishing point in
an image. In addition, some photos may solely focus on the objects (e.g., a flower) and do not possess a well-defined perspective geometry. We will analyze the composition of these images by separating the foreground objects from the background. Finally, apart from providing on-site feedback to photographers, our method can also be implemented as a component in large-scale image retrieval engines. One can develop a method to select only images with a dominant vanishing point to prepare the database off-line for on-line composition-sensitive retrieval requests.

There are also several directions we can explore to help portrait photographers design and analyze compositions. The relationship between triangles and the aesthetic quality of compositions can be further studied. For instance, how do the number, sizes, shapes, and orientations of triangles influence the aesthetics of photo composition? Answering this question can help amateur photographers learn more specific photography techniques. Moreover, our system can further assist users in shooting photos. Real-time suggestions about pose adjustments can be provided to users to help them embed more appealing triangles in their works.
Bibliography

[1] Pablo Arbelaez. Boundary extraction in natural images using ultrametric contour maps. In *Proceedings of Computer Vision and Pattern Recognition Workshop*, pages 182–182, 2006.

[2] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):898–916, 2011.

[3] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):898–916, 2011.

[4] Olga Barinova, Vadim Konushin, Anton Yakubenko, KeeChang Lee, Hwasup Lim, and Anton Konushin. Fast automatic single-view 3D reconstruction of urban scenes. In *Proceedings of the 10th European Conference on Computer Vision*, pages 100–113, 2008.

[5] Vasileios Belagiannis, Sikandar Amin, Mykhaylo Andriluka, Bernt Schiele, Nassir Navab, and Slobodan Ilic. 3D pictorial structures for multiple human pose estimation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1669–1676. IEEE, 2014.
[6] Subhabrata Bhattacharya, Rahul Sukthankar, and Mubarak Shah. A framework for photo-quality assessment and enhancement based on visual aesthetics. In *Proceedings of the ACM International Conference on Multimedia*, pages 271–280, 2010.

[7] Anjan Chatterjee. Neuroaesthetics: A coming of age story. *Journal of Cognitive Neuroscience*, 23(1):53–62, 2011.

[8] Di Dio Cinzia and Gallese Vittorio. Neuroaesthetics: A review. *Current Opinion in Neurobiology*, 19(6):682–687, 2009.

[9] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Wang. Studying aesthetics in photographic images using a computational approach. In *Proceedings of 2006 European Conference on Computer Vision*, pages 288–301. Springer, 2006.

[10] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. Studying aesthetics in photographic images using a computational approach. In *Proceedings of the 9th European Conference on Computer Vision*, pages 288–301, 2006.

[11] Ritendra Datta, Jia Li, and James Z Wang. Learning the consensus on visual quality for next-generation image management. In *Proceedings of the 15th ACM International Conference on Multimedia*, pages 533–536, 2007.

[12] Sagnik Dhar, Vicente Ordonez, and Tamara L. Berg. High level describable attributes for predicting aesthetics and interestingness. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1657–1664, 2011.
[13] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[14] Vittorio Ferrari, Manuel Marin-Jimenez, and Andrew Zisserman. Progressive search space reduction for human pose estimation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[15] Georgia Gkioxari, Bharath Hariharan, Ross Girshick, and Jitendra Malik. Using k-poselets for detecting people and localizing their keypoints. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 3582–3589. IEEE, 2014.

[16] Stephen Gould, Richard Fulton, and Daphne Koller. Decomposing a scene into geometric and semantically consistent regions. In *Proceedings of the IEEE 12th International Conference on Computer Vision*, pages 1–8, 2009.

[17] Abhinav Gupta, Alexei A Efros, and Martial Hebert. Blocks world revisited: Image understanding using qualitative geometry and mechanics. In *Proceedings of the 11th European Conference on Computer Vision*, pages 482–496, 2010.

[18] Feng Han and Song-Chun Zhu. Bottom-up/top-down image parsing by attribute graph grammar. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):59–73, 2009.

[19] Varsha Hedau, Derek Hoiem, and David Forsyth. Thinking inside the box: Using appearance models and context based on room geometry. In *Proceedings of the 11th European Conference on Computer Vision*, pages 224–237, 2010.
[20] Varsha Hedau, Derek Hoiem, and David A. Forsyth. Recovering the spatial layout of cluttered rooms. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, pages 1849–1856, 2009.

[21] Derek Hoiem, Alexei A. Efros, and Martial Hebert. Automatic photo pop-up. ACM Transactions on Graphics, 24(3):577–584, 2005.

[22] Derek Hoiem, Alexei A Efros, and Martial Hebert. Recovering surface layout from an image. International Journal of Computer Vision, 75(1):151–172, 2007.

[23] Thomas Jacobsen. Bridging the arts and sciences: A framework for the psychology of aesthetics. Leonardo, 39(2):155–162, 2006.

[24] Xin Jin, Mingtian Zhao, Xiaowu Chen, Qinping Zhao, and Song-Chun Zhu. Learning artistic lighting template from portrait photographs. In Proceedings of the 11th European Conference on Computer vision, pages 101–114, 2010.

[25] Dhiraj Joshi, Ritendra Datta, Elena Fedorovskaya, Quang-Tuan Luong, James Z Wang, Jia Li, and Jiebo Luo. Aesthetics and emotions in images. IEEE Signal Processing Magazine, 28(5):94–115, 2011.

[26] Yan Ke, Xiaou Tang, and Feng Jing. The design of high-level features for photo quality assessment. In Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, pages 419–426, 2006.
[27] Hui Kong, Jean-Yves Audibert, and Jean Ponce. Vanishing point detection for road detection. In *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 96–103, 2009.

[28] Jana Kosecká and Wei Zhang. Video compass. In *Proceedings of the 7th European Conference on Computer Vision*, pages 476–490, 2002.

[29] David A. Lauer and Stephen Pentak. *Design Basics*. Cengage Learning, 2011.

[30] Helmut Leder, Benno Belke, Andries Oeberst, and Dorothee Augustin. A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, 95(4):489–508, 2004.

[31] David C. Lee, Martial Hebert, and Takeo Kanade. Geometric reasoning for single image structure recovery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2136–2143, 2009.

[32] Jia Li. Agglomerative connectivity constrained clustering for image segmentation. *Statistical Analysis and Data Mining*, 4(1):84–99, 2011.

[33] Ligang Liu, Renjie Chen, Lior Wolf, and Daniel Cohen-Or. Optimizing photo composition. *Computer Graphics Forum*, 29(2):469–478, 2010.

[34] Paul Locher. Contemporary experimental aesthetics: State of the art technology. *i-Perception*, 2(7):697, 2011.
[35] Wei Luo, Xiaogang Wang, and Xiaoou Tang. Content-based photo quality assessment. In *Proceedings of 2011 IEEE International Conference on Computer Vision*, pages 2206–2213, 2011.

[36] Yiwen Luo and Xiaoou Tang. Photo and video quality evaluation: Focusing on the subject. In *Proceedings of the 10th European Conference on Computer Vision*, pages 386–399, 2008.

[37] Luca Marchesotti, Florent Perronnin, Diane Larlus, and Gabriela Csurka. Assessing the aesthetic quality of photographs using generic image descriptors. In *Proceedings of 2011 IEEE International Conference on Computer Vision*, pages 1784–1791, 2011.

[38] Hossein Mobahi, Shankar Rao, Allen Y. Yang, Shankar S. Sastry, and Yi Ma. Segmentation of natural images by texture and boundary compression. *International Journal of Computer Vision*, 95(1):86–98, 2011.

[39] Vladimir Nedovic, Arnold W. M. Smeulders, André Redert, and Jan-Mark Geusebroek. Stages as models of scene geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1673–1687, 2010.

[40] Masashi Nishiyama, Takahiro Okabe, Imari Sato, and Yoichi Sato. Aesthetic quality classification of photographs based on color harmony. In *Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition*, pages 33–40, 2011.
[41] Pere Obrador, Rodrigo de Oliveira, and Nuria Oliver. Supporting personal photo storytelling for social albums. In *Proceedings of the ACM International Conference on Multimedia*, pages 561–570, 2010.

[42] Pere Obrador, Ludwig Schmidt-Hackenberg, and Nuria Oliver. The role of image composition in image aesthetics. In *Proceedings of the 14th IEEE International Conference on Image Processing*, pages 3185–3188, 2010.

[43] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.

[44] Razvan Orendovici and James Z Wang. Training data collection system for a learning-based photographic aesthetic quality inference engine. In *Proceedings of the ACM International Conference on Multimedia*, pages 1575–1578, 2010.

[45] Gabriele Peters. Aesthetic primitives of images for visualization. In *Proceedings of 11th International Conference on Information Visualization*, pages 316–325. IEEE, 2007.

[46] Vilayanur S Ramachandran and William Hirstein. The science of art: A neurological theory of aesthetic experience. *Journal of Consciousness Studies*, 6(6-7):6–7, 1999.

[47] William M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971.
[48] Christopher Rasmussen. Grouping dominant orientations for ill-structured road following. In *Proceedings of the 2004 IEEE Conference on Computer Vision and Pattern Recognition*, pages 470–477, 2004.

[49] Rolf Reber, Norbert Schwarz, and Piotr Winkielman. Processing fluency and aesthetic pleasure: is beauty in the perceiver’s processing experience? *Personality and Social Psychology Review*, 8(4):364–382, 2004.

[50] Miriam Redi and Bernard Merialdo. Enhancing semantic features with compositional analysis for scene recognition. In *Proceedings of the 12th International Conference on Computer Vision*, pages 446–455, 2012.

[51] Bryan C. Russell, Alexei A. Efros, Josef Sivic, Bill Freeman, and Andrew Zisserman. Segmenting scenes by matching image composites. In *Advances in Neural Information Processing Systems*, pages 1580–1588, 2009.

[52] Benjamin Sapp and Ben Taskar. Modec: Multimodal decomposable models for human pose estimation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 3674–3681. IEEE, 2013.

[53] Ashutosh Saxena, Min Sun, and Andrew Y. Ng. Make3D: Learning 3D scene structure from a single still image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):824–840, 2009.

[54] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
[55] Peter D Stebbing. A universal grammar for visual composition? *Leonardo*, 37(1):63–70, 2004.

[56] Hsiao-Hang Su, Tse-Wei Chen, Chieh-Chi Kao, Winston H Hsu, and Shao-Yi Chien. Scenic photo quality assessment with bag of aesthetics-preserving features. In *Proceedings of the 19th ACM International Conference on Multimedia*, pages 1213–1216, 2011.

[57] Jean-Philippe Tardif. Non-iterative approach for fast and accurate vanishing point detection. In *Proceedings of the IEEE 12th International Conference on Computer Vision*, pages 1250–1257, 2009.

[58] Ryan Tokola, Wongun Choi, and Silvio Savarese. Breaking the chain: liberation from the temporal markov assumption for tracking human poses. In *Proceedings of IEEE International Conference on Computer Vision*, pages 2424–2431. IEEE, 2013.

[59] Alexander Toshev and Christian Szegedy. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1653–1660. IEEE, 2014.

[60] Elena Tretiak, Olga Barinova, Pushmeet Kohli, and Victor S. Lempitsky. Geometric image parsing in man-made environments. *International Journal of Computer Vision*, 97(3):305–321, 2012.

[61] Roberto Valenzuela. *Picture Perfect Practice: A Self-Training Guide to Mastering the Challenges of Taking World-Class Photographs*. New Riders, 2012.
[62] R Grompone Von Gioi, Jeremie Jakubowicz, Jean-Michel Morel, and Gregory Randall. LSD: A line segment detector. *Image Processing On Line*, 2(3):5, 2012.

[63] Chunyu Wang, Yizhou Wang, Zhouchen Lin, Alan L Yuille, and Wen Gao. Robust Estimation of 3D Human Poses from a Single Image. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2369–2376. IEEE, 2014.

[64] Yi Yang and Deva Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1385–1392. IEEE, 2011.

[65] Lei Yao, Poonam Suryanarayan, Mu Qiao, James Z. Wang, and Jia Li. OSCAR: On-site composition and aesthetics feedback through exemplars for photographers. *International Journal of Computer Vision*, 96(3):353–383, 2012.

[66] Yanhao Zhang, Xiaoshuai Sun, Hongxun Yao, Lei Qin, and Qingming Huang. Aesthetic composition representation for portrait photographing recommendation. In *Proceedings of the 19th IEEE International Conference on Image Processing*, pages 2753–2756, 2012.

[67] Silvia Zuffi, Javier Romero, Cordelia Schmid, and Michael J Black. Estimating human pose with flowing puppets. In *Proceedings of IEEE International Conference on Computer Vision*, pages 3312–3319. IEEE, 2013.