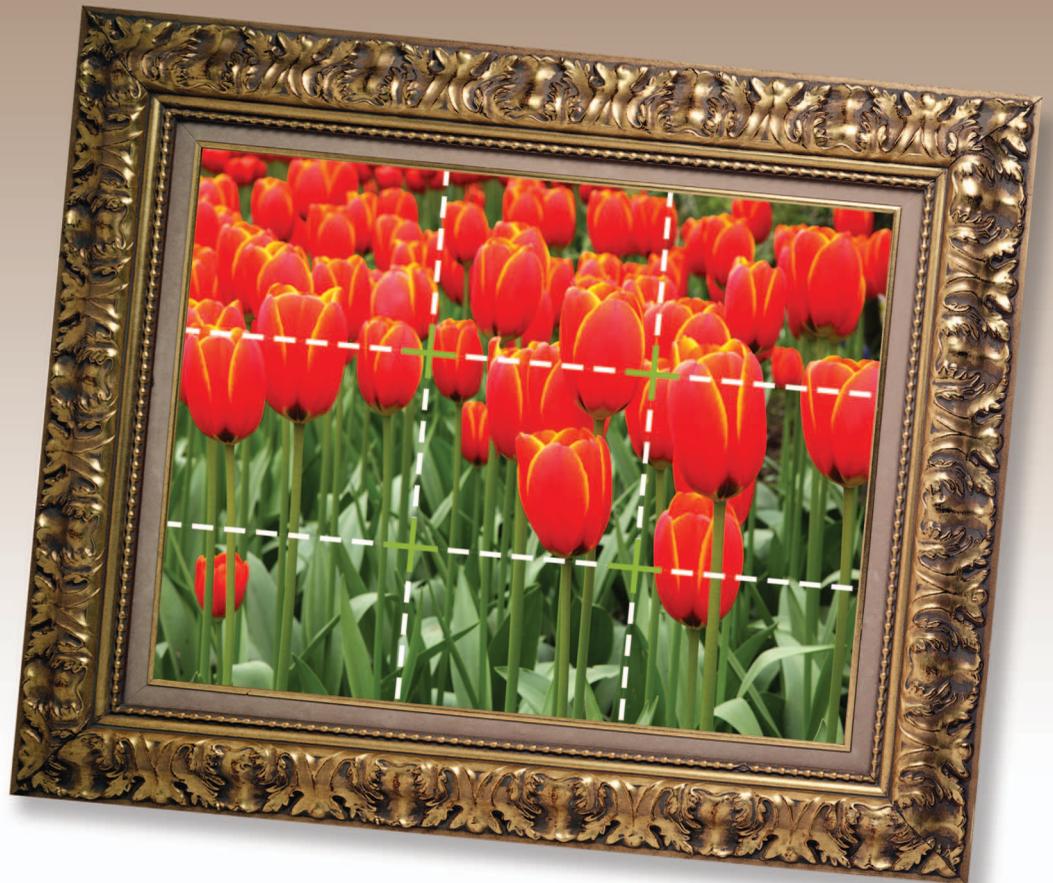


Dhiraj Joshi,
Ritendra Datta,
Elena Fedorovskaya,
Quang-Tuan Luong,
James Z. Wang,
Jia Li, and Jiebo Luo



PUBLICDOMAINPICTURES.NET &
© BRAND X PICTURES

Aesthetics and Emotions in Images

[A computational perspective]

In this tutorial, we define and discuss key aspects of the problem of computational inference of aesthetics and emotion from images. We begin with a background discussion on philosophy, photography, paintings, visual arts, and psychology. This is followed by introduction of a set of key computational problems that the research community has been striving to solve and the computational framework required for solving them. We also describe data sets available for performing assessment and outline several real-world applications where research in this domain can be employed. A significant number of papers that have attempted to solve problems in aesthetics and emotion inference are surveyed in this tutorial. We also discuss future directions that researchers can pursue and make a strong case for seriously attempting to solve problems in this research domain.

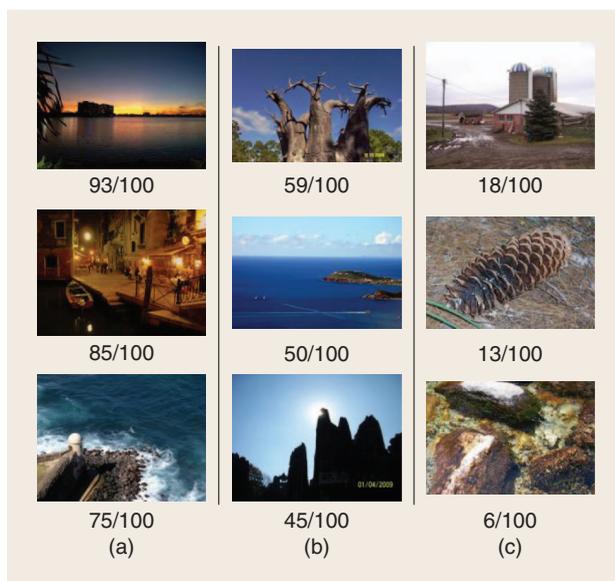
Digital Object Identifier 10.1109/MSP.2011.941851
Date of publication: 22 August 2011

INTRODUCTION

The image processing community together with vision and computer scientists have, for a long time, attempted to solve image quality assessment [78], [47], [18], [96] and image semantics inference [20]. While the former deals primarily with the quantification of low-level perceptual degradation of an image (typically from its original version), the latter attempts to infer the content of an image and associate high-level semantics to it, in part or in whole. More recently, researchers have drawn ideas from the aforementioned to address yet more challenging problems such as associating pictures with aesthetics and emotions that they arouse in humans, with low-level image composition [19], [21], [91], [92]. Because emotions and aesthetics also bear high-level semantics, it is not a surprise that research in these areas is heavily intertwined. Besides, researchers in aesthetic quality inference also need to understand and consider human subjectivity and the context in which the emotion or aesthetics is perceived. As a result, ties between computational image analysis and psychology, study of beauty [54], [70], and aesthetics in visual art, including photography, are also natural and essential. The key challenges for researchers are the loose and highly subjective nature of semantics associated with emotions and aesthetics, and the seemingly inherent semantic gap between low-level computable visual features and high-level human-oriented semantics.

Despite the challenges, various research attempts have been made and are increasingly being made to address basic understanding and solve various subproblems under the umbrella of aesthetics, mood, and emotion inference in pictures. What motivates the multidisciplinary community to make such attempts is the fact that there is much to be gained from systems that can indeed reliably infer, at least for a section of the population, what the perceptual, cognitive, aesthetic, and emotional response to a photograph or a visual artwork will be. The potential beneficiaries of this research include general consumers, media management vendors, photographers, and people who work with art. Good shots or photo opportunities may be recommended to consumers; media personnel can be assisted with good images for illustration while interior and healthcare designers can be helped with more appropriate visual design items.

Given that many Web-based image repositories (e.g., Flickr) are currently multibillion images in size, semantics can no longer be the sole criterion for image search and organization. Moreover, many image hosting and sharing Web sites have recognized the need for introducing some form of aesthetic or appeal measure (discussed in the section “Key Problems in Aesthetics and Emotions Inference”). Aesthetic appeal can help find exciting and appealing photographs from large collections while sorting out unappealing ones. Such a system can also be embedded into digital cameras to provide live feedback on the potential visual appeal of a shot at a given time, or software that can help a user to aesthetically design albums, slide shows, and other photo related products (discussed in the section “Computational Frameworks”).



[FIG1] Pictures with (a) high, (b) medium, and (c) low aesthetics scores from the Aesthetic Quality Inference Engine (ACQUINE).

Editing is a key element in professional photography. Picture editors (including photographers) usually must review a large collection of images to select the strongest ones for different photographic causes. The nature of selection could vary with the nature of the cause (e.g., selection can be different for a photographer's participation in photography rating sites, photo-clubs, competition, portfolio reviews, or photography workshops). An automated system that can provide feedback about aesthetics or quality based on learned rules could be a very useful aid in picture editing (Figure 1 shows an example of state-of-the-art automatic aesthetics assessment). Similarly, for a museum curator, considerations about a piece of art are about its importance, its relevance as a good example of the current concerns of artists and society, its originality, and freshness in content. From a publication perspective, a curator may be interested in assessing if an artwork is enjoyable by a majority of the people. Again, we see that based on the need or the cause for determining the goodness of fit of an artwork, an expert curator may use different judgment built by years of experience. An art historian usually analyzes a body of work of one or several artists to make inferences. Techniques that study similarities and differences between artists and artwork at the aesthetic level could be of value to art historians. We strongly believe that computational models of aesthetics and emotions may be able to assist in this decision making and perhaps with time and feedback learn to adapt to expert opinion better (Figure 2 shows user-rated emotions under the framework of Web image search that can potentially be used for learning emotional models). Computational aesthetics does not intend to obviate the need for expert opinion. On the other hand, automated methods would strive toward becoming useful suggestion systems for experts that can be personalized (to

one or few experts) and improved with feedback over time (as also expressed in [84]).

It is widely believed by artists that the most important element about a work of art is not aesthetics but the idea behind its conception. By art, we refer to any one of the forms of creative art (discussed in the section “Background”). Therefore, for art, aesthetics is a derived quality that is intertwined with genre, context, and semantics of the artwork (e.g., the qualities that make a wedding picture aesthetically beautiful are different from those that make an aesthetically good picture of a church, although the two pictures may be taken at the same location). Therefore the aesthetics of an artwork may also vary with choice of subject. Another important criterion that influences perception of aesthetics is the level of sophistication of the viewer. Artists generally have different perception of aesthetics than do less knowledgeable viewers of art. A great piece of art may at times be perceived as boring by a less knowledgeable viewer. Perception of aesthetics of an artwork is therefore at least a complex function of the artwork, the intent of the artist, the semantics conveyed and perceived, the genre of the art, and the level of experience of the viewer. A true scientific analysis would entail a controlled individual variation of the above factors to determine their effects on perceived aesthetics. However, in reality, some of the above factors are difficult to determine and more so to vary (e.g., intent of an artist or genre of the art). Therefore, the more realistic goal of computational methods is probably to aim for the universally adopted aspects of aesthetics and emotion evoked by commonly seen subjects.

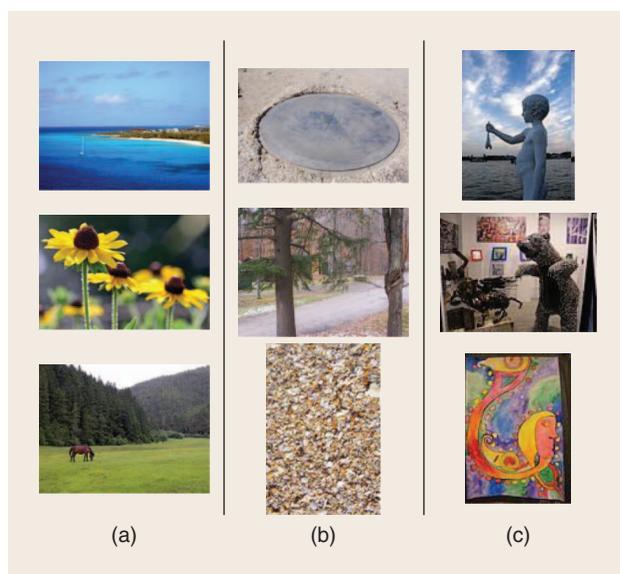
IT IS WIDELY BELIEVED BY ARTISTS THAT THE MOST IMPORTANT ELEMENT ABOUT A WORK OF ART IS NOT AESTHETICS BUT THE IDEA BEHIND ITS CONCEPTION.

In statistical theory of estimation, an estimator of a parameter θ is denoted as $\hat{\theta}$ where the estimator asymptotically approaches the true value of the parameter when the sample size grows to infinity. In a similar analogy, the

true aesthetic value (or true aesthetics distribution in case of a lack of singular consensus even among experts) is perhaps identifiable only when the sample size is infinitely large and there is no noise in observations. We would like to differentiate between aesthetics of an artwork—the true aesthetics value or distribution (an asymptotic concept perhaps determinable by experts (such as the artists themselves) whose experience tends to infinity) and the observed aesthetics of an artwork—the aesthetics obtained from a pool of values determined by a mix of experts and general viewers. We would like to qualify that the study of computational aesthetics can at best attempt to model observed aesthetics under a constrained population. While such a modeling is limited within the true scope and definition of aesthetics that has pervaded art over centuries, we believe that it is an earnest and developing attempt to explain abstract phenomena.

While the current area of computational visual aesthetics may still not be much beyond its infancy, community interest can be gauged by the active attendance and participation in recent related forums, including the special session on Image Aesthetics, Moods, and Emotions in the 2008 IEEE International Conference on Image Processing, which was cochaired by three of the coauthors. In parallel, a special session on Art and Perception has been a regular part of IS&T/SPIE Electronic Imaging Conference through several years, cochaired by one of the coauthors. In the past, the signal processing community has devoted special issues to young and challenging research areas [106], [107], [108]. While tutorials are typically written for relatively mature topics, we believe an early tutorial on this active topic will help summarize the existing attempts, conjure up future research directions, and ultimately lead to robust solutions.

Computational methods have for decades attempted to impose orders or constraints on models to explain the observed phenomena. All scientific theories are built upon certain premises or assumptions at their foundation (sometimes unverifiable). Certain premises or theories are disproven or corrected with time to give way to improved theories but such are the ways of science and every honest attempt counts toward pushing the scientific frontier. In this survey, we discuss research that attempts to explain the observed phenomena of aesthetics and emotions that arise from subjective judgments using known tools and knowledge about computer vision, machine learning, art, and photography (sections “Key Problems in Aesthetics and Emotions Inference” and “Computational Frameworks”). This article attempts to pave the way for increased participation between domain experts in different fields of art and computer scientists.



[FIG2] Pictures and emotions rated by users from ALIPR.com, a research site for machine-assisted image tagging: (a) pleasing, (b) boring, and (c) surprising.

**THIS ARTICLE ATTEMPTS TO
PAVE THE WAY FOR INCREASED
PARTICIPATION BETWEEN DOMAIN
EXPERTS IN DIFFERENT FIELDS OF ART
AND COMPUTER SCIENTISTS.**

In this tutorial, we have attempted to introduce components that are essential for the broader research community to get involved and excited about this field of study. It is our hope that this tutorial will attempt to tie the related areas of semantics inference, image aesthetics, and emotions together and draw useful links with research in philosophy, psychology, and visual arts.

BACKGROUND

The word “aesthetics” originates from the Greek word *aisthētikos* sensitive, derived from *aisthanesthai* “to perceive, to feel.” The *American Heritage Dictionary of the English Language* provides the following currently used definitions of aesthetics:

- 1) the branch of philosophy that deals with the nature and expression of beauty, as in the fine arts. In Kantian philosophy, the branch of metaphysics concerned with the laws of perception
- 2) the study of the psychological responses to beauty and artistic experiences
- 3) a conception of what is artistically valid or beautiful
- 4) an artistically beautiful or pleasing appearance.

Philosophical studies in aesthetics (as well as the philosophy of art) focus on questions such as “What is beautiful (ugly) in nature and in art,” “What are the principles of aesthetic judgments,” “What constitutes a work of art,” “How beauty and art relate to truth,” “How art can be interpreted and evaluated,” and “What states of mind—perceptions, attitudes, and emotions—are involved in aesthetic experience.” Many of these questions were originally proposed by Plato, and further developed in the works of Aristotle, Hutcheson, Baumgarten, Hume, Kant, and others (see *Encyclopedia Britannica* [112]), leading to the formulation of two traditional views on beauty and aesthetic values. The first view considers aesthetic values to be objectively existing and universal, while the second position treats beauty as a subjective phenomenon, depending on the attitude of the observer.

Contemporary scholars in philosophy and humanities have advanced our views of aesthetics, particularly with respect to pictorial representation and emphasized “meaning” as a chief determinant of aesthetics. According to Goodman [39], artworks are composed of symbols that refer to the worlds we construct. Therefore, understanding art and its aesthetic principles requires cognitive interpretation of these symbols. Such an interpretation depends to a large extent on what is familiar and habitual in the existing cultural environment and on the syntactical and semantic rules that are used in the process of referencing. The mode of referencing utilized in pictures is denotation—pictures are labels for the world of our experience categories. To aesthetically evaluate pictures, artistic symbols are to be uncovered and judged based on principles similar to those existing in other domains of human knowledge, e.g., science, what new understanding of

the world these symbols and corresponding classifications bring about, how they influence our perception and relationship to the world, and what emotions they evoke. Goodman suggested several symptoms of the

aesthetic—characteristics of symbol systems occurring in art. Another contemporary scholar in humanities, Mitchell has extensively studied the relations of visual and verbal representations in art. He has noted that the interconnections between writing and depiction define the aesthetic value of pictorial art [66], [67]. Unlike Goodman, Wollheim emphasized resemblance and also stressed the importance of psychological context and artistic intention to uncover depictive meaning [99]. He argued that there exists a standard of correctness for pictorial representation, which is necessary to evoke intended feelings. Elkins, an eminent scholar in art history and aesthetics, has proposed a two-level model of depiction; the first level corresponds to resemblance between depictions and objects and relations, and the second level corresponds to rules of interpretation, thus adding notion of resemblance to Goodman’s aesthetics [26]. In a recent paper [27], Elkins presents an argument that art and science do not really have sufficiently developed common ground with respect to aesthetics despite existing attempts to unify their ideas and approaches on aesthetics. At the same time, certain scholars in both sciences and humanities believe that common ground can potentially be found in the fields of evolutionary psychology and cognitive science (e.g., Dutton [25]).

A PERSPECTIVE ON PHOTOGRAPHS

While aesthetics can be colloquially interpreted as a seemingly simple matter as to what is beautiful, few can meaningfully articulate the definition of aesthetics or how to achieve a high level of aesthetic quality in photographs. There is still a need to have a formal or mathematical explanation of aesthetics in photographs. It is widely believed and can often be experimentally demonstrated that aesthetics is at times very subjective. That is, the same photograph can be appreciated by some viewers but not by certain others. The “taste” and sophistication of the viewer often determines the aesthetic rating given by the viewer. For years, Photo.net has been a place for photographers to rate the photos of peers [115]. Here a photo is rated along two dimensions, aesthetics and originality, each with a score between one and seven. In terms of aesthetics, Photo.net explains that the score of one means ugly and the score of seven means beautiful. Example reasons for a high rating include “looks good, attracts/holds attention, interesting composition, great use of color, (if photo journalism) drama, humor, and impact, and (if sports) peak moment, struggle of athlete.” The formation of a person’s aesthetic opinion can be subtle. It certainly involves more than what has been genetically coded during our millions of years of evolution. For instance, social background can be critical when we look at photographs involving human activities. The

knowledge of a person helps to understand the intrinsic meanings, the cultural implications, the emotional resonance, or the values expressed through the arrays of pixels. The personal attitude and instinct can also be determining factors. Further, because these factors are all dynamic, a person's aesthetic opinion can change over time and over context.

Ideas of aesthetics emerged in photography around the late 19th century with a movement called Pictorialism. Because photography was a relatively new art at that time, the Pictorialist photographers drew inspiration from paintings and etchings to the extent of emulating them directly. The goal was to shoot and develop pictures that had artistic quality (as perhaps a predictor of aesthetics). Photographers used techniques such as soft focus, special filters, lens coatings, special darkroom processing, and printing to achieve desired effects. By around 1915, the widespread cultural movement of Modernism had begun to affect the photographic circles. In Modernism, ideas such as formal purity, medium specificity, and originality of art became paramount. Modernism created a divide between low art and high art by emphasizing formal aesthetic qualities of art. Post-Modernism rejected ideas of objective truth in art. Sharp classifications into high art and low art became defunct. A post-Modernist artist did not conform to any forms of dualisms nor rigid genre boundaries.

In spite of these factors, certain patterns stand out with respect to photographic aesthetics. This is especially true in certain domains of photography. For example, in nature photography, it can be demonstrated that the appreciation of striking scenery is universal. Nature photographers often share common techniques or rules of thumb in their choices of colors, tonality, lighting, focus, content, vantage point, and composition. As one example, to impress the viewers, nature landscape photographers often prefer to use one type of slide film, the Fuji Velvia film, even though it is well known that the film produces very saturated and high-contrast photos rather than capturing the true colors of the real world. The purer the primary colors, red (sunset, flowers), green (trees, grass), and blue (sky), the more striking the scenery is to viewers. In terms of composition, there are common and not-so-common theories or rules. The rule of thirds is the most widely known. It basically states that the most important part of the image is not the exact center of the image but rather at the one third and two third lines (both horizontal and vertical), and their four intersections. The viewer's eyes can naturally concentrate on these areas than either the center or the borders of the image. In composition, it is often beneficial to place objects of interest in these areas. A less common rule in nature photography is to use diagonal lines (such as a railway, a line of trees, a river, or a trail) or converging lines for the main objects of interest to draw the attention of the human eyes. Another composition rule is to frame the shot so that there are interesting objects in both the close-up foreground and the far-away background. For example, when shooting a photo of a moun-

tain range, it is often better to add some foreground objects such as trees, flowers, or animals. However, great photographers often have the talents to know when to break these rules to be more creative. Ansel Adams said, "There are no rules for good photographs, there are only good photographs." Another renowned American photographer, Edward Weston, said, "To consult the rules of composition before making a picture is a little like consulting the law of gravitation before going for a walk. Such rules and laws are deduced from the accomplished fact; they are the products of reflection." Given how complex this problem is, it is of course too early to expect a computer program to be able to infer aesthetic quality of photographs in the same fashion as humans do. Such a computer program must be an enormous knowledge engine that can comprehend many of the objects in the world as well as understand the perceptions of our human society.

A PERSPECTIVE ON PAINTINGS

Painters in general have a much greater freedom to play with the palette, the canvas, and the brush to capture the world and its various seasons, cultures, and moods. Techniques of drawing and painting that assure great accuracy in the depiction of the real world do not always guarantee astounding beauty. In some more extreme opinions, copying nature can be thought of as the work of a technician rather than an artist. Photographs at large represent true physical constructs of nature (although film photographers sometimes aesthetically enhanced their photos by dodging and burning). Artists, on the other hand, have always used nature as a base or as a "teacher" to create works that reflected their feelings, emotions, and beliefs. Although many artists suggest that beginners should never stop learning from nature, they often also stress the representative aspect of painting or drawing, which is totally different from photography.

History abounds with many influential art movements that dominated the world art scene for certain periods of time and then faded away, making room for newer ideas. It would not be incorrect to say that most art movements (sometimes individual artists) defined characteristic painting styles that became the primary determinants of art aesthetics of the time. At times, the influence of painters or pre-eminence of styles were recognized posthumously or in retrospect.

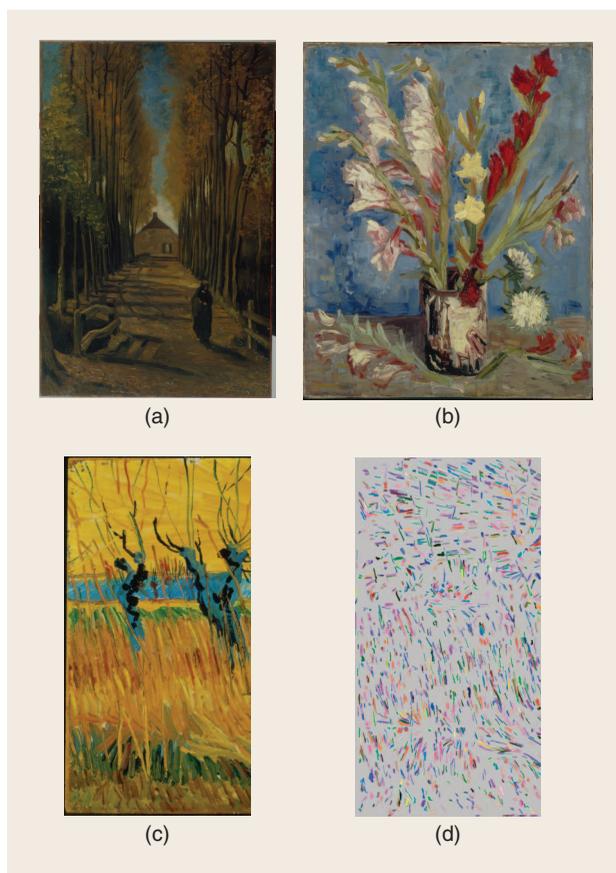
In painting as in photography, aesthetics evolves as bold ideas are introduced, practiced, and accepted by the world. One of the key turning points in Western art occurred in late 19th century when a few radical Parisian painters decided to hold their own art exhibition as a rebellion against traditional studio painting. Impressionism (the movement that followed), derived its name from Claude Monet's masterpiece *Impression, Sunrise, 1872*. Impressionist artists focused on ordinary subject matter, painted outdoors, used visible brushstrokes, and employed colors to emphasize light and its effect on their subjects. A derivative movement, Pointillism, was pioneered by Georges Seurat, who mastered the art of using

colored dots as building blocks for paintings. Pointillism presented a fresh approach to mixing of colors wherein additive mixing of primary pigments was performed by human eye (on seeing the colored dots) as opposed to traditional mixing of colors in the palette (which is a subtractive mixing), often giving a vibrant look to paintings.

Early 20th century post-Impressionist artists digressed from the past and introduced a personal touch to their world depictions giving expressive effects to their paintings. Van Gogh is especially known for his bold and forceful use of colors to express his artistic ideas (Figure 3). His use of color varied over time and was often a deep reflection of the nature of his subjects, his interactions with other artists, and his own emotions. Van Gogh also developed a bold style of brushstrokes, an understanding of which can perhaps offer newer perspectives into understanding his work and that of his contemporaries (Figure 3 shows an example of automatic brushstroke extraction research presented in [45]).

With the rise of expressionism, blending of reality and artists' emotions became vogue. Expressionist artists freely distorted reality into a personal emotional expression. Abstract expressionism, a post-World War II phenomenon, put America in the center stage of art for the first time in history. Intense personal expression combined with spontaneity and hints of subconscious and surreal emotion gave a strikingly new meaning to art. Possibilities of creation became virtually unbounded. A recent work [88] scientifically examines the works of Mondrian and Pollock, two stalwarts of Modern art with drastically distinct styles. Mondrian, a veteran European-American artist, believed in spiritual harmony in art. He strived to achieve an aesthetic balance in his compositions through appropriate arrangements of lines, surfaces, and use of primary colors. Mondrian's painting style came to be known as Neoplasticism and was in some forms inspired from earlier movements such as Pointillism (Degas's style) and Cubism (Picasso's style). In contrast to the careful and harmonic style of Mondrian, Jackson Pollock professed a shockingly unconventional painting technique moving away from the easel and the brush, spreading his canvas on the floor, using hardened brushes or sticks to paint, and dripping paint onto the canvas (known as the "drip technique"). He would sometimes mix sand or broken glass with paint to add texture to his artwork. Pollock's style drew both praise and disapproval from critics. In particular, Clement Greenberg, an eminent art critic, has termed Jackson Pollock's style as the epitome of aesthetic value in art [40]. In [87] and [88], physicists have attempted to explain the aesthetic significance of Pollock's art using fractal patterns that abound in nature and are also believed to be aesthetically pleasing to the eye.

While the relation of aesthetics and art is very intriguing and open to philosophical discussion, computational methods have made attempts to work on subproblems of the whole. In recent years, as artistic paintings are digitized in museums and galleries with high-quality equipment, it is



[FIG3] Van Gogh's paintings (a) *Avenue of Poplars in Autumn*, (b) *Still Life: Vase with Gladioli*, (c) *Willows at Sunset*, and (d) automatically extracted brushstrokes for *Willows at Sunset*. Notice the widely different nature and use of colors in the paintings. Parts (a) and (b) are courtesy of the Van Gogh Museum Amsterdam (Vincent van Gogh Foundation). Parts (c) and (d) are courtesy of the Kröller-Müller Museum and James Z. Wang Research Group at Penn State.

becoming possible to study paintings using computational techniques. Existing work mainly focuses on several key issues: retrieval of similar paintings, authentication of painters, distinguishing painting styles, dating of paintings, and reconstruction of an original scene in 3-D. Although there has recently been work on inferring aesthetics in paintings [57], [87], [88], such work is limited to a small-scale specific experimental setup.

A PERSPECTIVE ON OTHER FORMS OF VISUAL ART

Beyond photography and paintings, aesthetics, emotion, and mood are essential to almost all artistic disciplines including sculpture, architecture, and crafts. In each of the ancient civilizations, Egypt, Mesopotamia, Greece, Rome, Persia, India, and China, many forms and styles of art were developed. The study of these other forms of visual art using computational means is further away from the scope of the signal processing community. Mathematical tools, such as geometry, are often indispensable in analyzing three-dimensional (3-D) objects.

With the assistance of computers, researchers have shown that by 1200 CE a conceptual breakthrough occurred in medieval Islamic architecture in which girih patterns were reconceived as tessellations of a special set of equilateral polygons decorated with lines, and these polygonal tiles enabled the creation of increasingly complex periodic girih patterns [61]. In another study, the curved spiral grooves carved on a class of ornamental jade burial rings from the spring and autumn period (771 to 475 BC) in China were analyzed using curve fitting [62]. Three-dimensional scanning and reconstruction of art objects have also been studied extensively. Since about 2000, researchers have used range and image sensing to create 3-D models of historical buildings [83]. Signal processing and geometry are used to match two-dimensional (2-D) curves and 3-D surfaces. For instance, Willis, and Cooper have used computers to assist the reconstruction of ancient artifacts such as broken ceramics [97].

AESTHETICS, EMOTIONS, AND PSYCHOLOGY

There are several main areas and directions of experimental research related to psychology, which focus on art and aesthetics: experimental aesthetics (psychology of aesthetics), psychology of art, and neuroaesthetics. These fields are interdisciplinary and draw on knowledge in other related disciplines and branches of psychology.

Experimental aesthetics is one of the oldest branches of experimental psychology, which officially begins with the publishing of Fechner's *Zur experimentalen Aesthetik* in 1871, and *Vorschule der Ästhetik* in 1876 [31], [32]. In his work, Fechner proposed a concept of bottom-up aesthetics. Fechner suggested three methods for use in experimental aesthetics, including the method of choice, where subjects are asked to compare objects with respect to their pleasingness; the method of production, where subjects are required to produce an object that conforms to their tastes by drawing or other actions; and the method of use, which analyzes works of art and other objects on the assumption that their common characteristics are those that are most approved in society. Fechner also adopted a wider concept of beauty by defining everything that had the property of immediately causing a liking as "beautiful," therefore making emotional response the central focus of his research.

Developments in other areas of psychology of the early decades of the 20th century contributed to the psychology of aesthetics. Gestalt psychology produced influential ideas such as the concept of goodness of configuration [51]. According to this school, we do not see isolated visual elements but instead patterns and configurations, which are formed according to the processes of perceptual organization in the nervous system (governed by the "law of Prägnanz"). This law enhances such properties as regularity, symmetry, simplicity, closure, and others, making us gravitate toward choosing "good" structures as preferred. Freud and other psychoanalysts have pursued the analogies between art and dream [35]. They analyzed the works of individual artists to show how the creation and appreciation of art can be explained as disguised expression for unfulfilled and repressed subconscious desires. The work of Rudolf

Arnheim, one of the most prominent authors in the psychology of art, was greatly influenced by the ideas of the Gestalt school in developing concepts related to balance, movement, shape, and representation of space [4]. He introduced terms of forces, strains, equilibrium, to discussing the principles of visual art. Another idea of Gestalt psychology relevant to aesthetics is called "physiognomics," which states that certain objects and human behavior are inherently expressive of specific emotional states. Thus, a weeping willow looks sad because willow branches convey the expression of passive hanging. In the 1970s, Berlyne revolutionized the field of experimental aesthetics by bringing to the forefront of the investigation psychophysiological factors and mechanisms underlying aesthetic behavior. In his seminal book *Aesthetics and Psychobiology* (1971) [9], Berlyne formulated several theoretically and experimentally substantiated ideas that helped shape modern experimental research in aesthetics into the science of aesthetics [69]. Berlyne noted that because art and aesthetic activity is apparently a feature of all of the 3,000 cultural forms on the earth's surface, it suggests that art grows out of some fundamental characteristics of the human nervous system.

Berlyne's ideas and research directions together with the advances in understanding of neural mechanisms of perception, cognition, and emotion obtained in psychology [82], psychophysiology, and neuroscience and facilitated by the modern imaging techniques led to the emergence of neuroaesthetics in the 1990s [46], [50], [72], [104]. According to Zeki, aesthetic sense corresponds to the specialized brain mechanisms (modules) that are involved in processing visual information, where those modules are tuned to analyze different aspects of visual images. It can thus be suggested that different art schools and artists could be selectively tuned or sensitized to emphasize the impression produced by the activity of certain brain mechanisms while painting. As an example, Zeki links the painting style of Mondrian and Malevich to the functioning of visual cortical areas of V_1 , V_2 , V_3 , and V_4 that are specialized to detect lines and their orientations, rectangular shapes, and colors. Analyzing why specific visual features evoke stronger aesthetic impressions than others, Latto concludes that a feature is intrinsically interesting if it resonates with the visual processing mechanisms [56]. Following these theories, Peters [71] proposes to consider the different modules of visual processing as the basic dimensions of visual aesthetics. Recent studies associated with the processing fluency theory by Reber et al. in [74] suggest that aesthetic experience is a function of the perceiver's processing dynamics: the more fluently the perceiver can process an image, the more positive is their aesthetic response. Ulrich and Gilpin [90] pointed out that numerous preference studies on visual environments including urban (architecture, interiors), as well as natural environments (forests, waterscapes) demonstrated a strong preferential tendency toward nature scenes compared to the urban scenes for population groups from different areas of the world. Evidence of a cross-cultural general people's dislike for abstract art and sculpture has been reported (based on polls conducted in many countries) in [100].

The previous work has received a mixed response and can be regarded as a cautionary story of democratic art taste.

KEY PROBLEMS IN AESTHETICS AND EMOTIONS INFERENCE

Many different problems have been studied under the umbrella of aesthetics and emotions evoked from pictures and paintings. While different problem formulations are focused on achieving different high-level goals, the underlying process is always aimed at modeling an appeal, aesthetics, or emotional response that a picture, a collection of pictures, or a piece of art evokes in human beings. Manifestations of appeal, aesthetics, or emotional responses could be (but are not limited to) appreciation of pictures, sentiment, taste, sensory attraction, appreciation of art, artistic characterization of paintings or pictures, photographic assessment of pictures, or simply mass appeal guided by important events and popular culture. Contrary to semantics, an aesthetics response is usually very subjective and difficult to gauge even among human beings. From a computational perspective, it is essential to discover ways to quantify these responses so as to mathematically formulate problems. User ratings provide a useful way to capture these values in a numeric form. Advantages of using population-driven response are that subjective patterns are captured as a whole while effects of outlier individual biases are toned down (“wisdom of the crowd”). Invariably, all formulations of aesthetics or emotional inference involve prediction of values from a discrete or continuous range. However the task at hand, the source and nature of data, the categories or prediction ranges, and the learning methodologies adopted can give different flavors to the problems.

We divide this discussion into two sections. The first section is devoted to mathematically formulating the core aesthetics and emotions prediction problems. In the second section, we discuss some problems that are directly or indirectly derived from the core aesthetics or emotions prediction problems in their scope or application. Here, we discuss problems that have seen a growth in research interest lately, while we do not claim to study an exhaustive list of associated problems in this article as the field is steadily evolving.

CORE PROBLEMS

AESTHETICS PREDICTION

When a photograph is rated by a set of n people on a 1-to- D scale on the basis of its aesthetics, the average score can be thought of as an estimator for its intrinsic aesthetic quality. More specifically, we assume that an image I has associated with it a true aesthetics measure $q(I)$, which is the asymptotic average if the entire population rated it. The average over the size n sample of ratings, given by $\hat{q}(I) = \frac{1}{n} \sum_{i=1}^n r_i(I)$ is an estimator for the population parameter $q(I)$, where $r_i(I)$ is the i th rating given to image I . Intuitively, a larger n gives a better estimate. A formulation for aesthetics score prediction is therefore to infer the value of $\hat{q}(I)$ by analyzing the content of image I , which is a direct emulation of humans in the photo rating process. This

lends itself naturally to a regression setting, whereby some abstractions of visual features act as predictor variables and the estimator for $\hat{q}(I)$ is the dependent variable. An attempt at regression-based score prediction has been reported in [19], showing limited success. The cited work assesses the quality of score prediction in the form of rate or distribution of error.

It has been observed both in [19] and [47] that score prediction is a highly challenging problem, mainly due to noise in user ratings. Given the limited size of rating samples, their averaged estimates have high variance, e.g., 5 and 5.5 on a one to seven scale could easily have been interchanged if a different set of users rated them, but there is no way to infer this from the content alone, which leads to large prediction errors. To make the problem more solvable, the regression problem is changed to one of classification, by thresholding the average scores to create high- versus low-quality image classes [19], or professional versus snapshot image classes [47]. Suppose threshold values are HIGH and LOW, respectively, then $\text{class}(I)$ is one if $\hat{q}(I) \geq \text{HIGH}$ and zero if $\hat{q}(I) \leq \text{LOW}$. When the band gap $\delta = \text{HIGH} - \text{LOW}$ increases, the two classes are more easily separable, a hypothesis that has been tested and found to hold in [19]. An easier problem, but one of practical significance, is that of selecting a few representative high-quality or highly aesthetic photographs from a large collection. In this case, it is important to ensure that most of the selected images are of high quality even though many of those not selected may be of high quality as well. An attempt at this problem [21] has proven to be more successful than the general HIGH/LOW classification problem described previously. The HIGH/LOW classification problem solutions can be evaluated by standard accuracy measures [19], [47]. Conversely, the selection of high-quality photos needs only to maximize the precision in high quality within the top few photos, with recall being less critical.

DISCUSSION

Prediction of aesthetics score is undoubtedly a finer form of prediction compared to prediction of a high/low aesthetics class. A score can potentially capture finer gradations of aesthetics values and hence a score predictor would be more valuable than an aesthetics class predictor. However, score prediction requires training examples from all spectrums of scores in the desired range and hence the learning problem is much more complex than the class prediction (which can typically be translated into a multiclass classification problem well known in machine learning). Another issue is the selection of an appropriate range of values for prediction. In short, prediction of the aesthetics class is a recourse that is taken to make aesthetics prediction more tractable. Finer scores are more desirable than high/low aesthetics values in general. Opportunities lie in i) exploring multiclass classification versus regression paradigms (discussed later) for score prediction, ii) building large and reliable data sets for learning, iii) performing psychological studies on people to understand how and in what scenarios humans perform class versus score prediction inside their brains, and iv) learning and predicting “distributions of aesthetics values” instead of

singular aesthetics classes or scores. The last problem is interesting in several regards; scores or values being ordinal rather than categorical in nature can be mapped to the real number space. Learning distribution of aesthetics on a per image basis can throw useful light on human perception and help algorithmically segment people into “perception categories.” Such research can also help characterize various gradations of “artist aesthetics” and “consumer aesthetics” and study how they influence one another perhaps over time. From a learning standpoint, a large data set labeled by a very diverse audience would be essential for such analysis. Modeling distribution of aesthetics can be approached within a multiclass classification framework where images are allowed to be classified into multiple classes. Knowledge about gradations or segments of rating population may further be leveraged for more precise modeling.

EMOTION PREDICTION

If we group emotions that natural images arouse into categories such as “pleasing,” “boring,” and “irritating,” then emotion prediction can be conceived as a multiclass categorization problem [102]. These categories are fuzzily defined and judgments are highly subjective. Consider that there are K such emotion categories, and people select one or more of these categories for each image. If an image I receives votes in the proportion, $\Pi_1(I), \dots, \Pi_K(I)$ then two possible questions arise.

MOST DOMINANT EMOTION

We wish to predict, for an image I , the most voted emotion category $k(I)$, i.e., $k(I) = \operatorname{argmax}_i \Pi_i(I)$. The problem is only meaningful when there is clear dominance of $k(I)$ over others, thus only these samples must be used for learning.

EMOTION DISTRIBUTION

Here, we wish to predict the distribution of votes (or an approximation) that an image receives from users, i.e., $\Pi_1(I), \dots, \Pi_K(I)$, which is well suited when images are fuzzily associated with multiple emotions.

The “most dominant emotion” problem is assessed like any standard multiclass classification problem. For “emotion distribution,” assessment requires a measure of similarity between discrete distributions, for which Kullback-Leibler (KL) divergence is a possible choice.

DISCUSSION

While the most dominant emotion prediction translates the problem into a multiclass classification problem that has successfully been attempted in machine learning, emotion distribution would be a more realistic and interesting problem from a human standpoint. Human beings rarely associate definitive emotions with pictures. In fact, it is believed that great works of art evoke a “mix of emotions” leaving little space for emotional purity, clarity, or consistency. However, learning a distribution of emotions from pictures requires a large and reliable emotion ground truth data set. At the same time, emotional categories are not completely independent (e.g.,

there may be correlations between “boring” and “irritating”). One of the key open issues in this problem is settling upon a set of plausible emotions that are experienced by human beings. Opportunities also lie in attempting to explore the relationships (both causal and semantic) between human emotions and leveraging them for prediction.

ASSOCIATED PROBLEMS

IMAGE APPEAL, INTERESTINGNESS, AND PERSONAL VALUE

Often, the appeal that a picture makes on a person or a group of people may depend on factors not easily describable by low-level features or even image content as a whole. Such factors could be sociocultural, demographic, purely personal (e.g., “a grandfather’s last picture”), or influenced by important events, vogues, fads, or popular culture (e.g., “a celebrity wedding picture”). In the age of ever-evolving social networks, “appeal” can also be thought of as being continually reinforced within a network framework. Facebook allows users to “like” a picture, a conversation, or a personal status or update, and it is not unusual to find “liking” patterns governed by one’s friends and network (e.g., a person is likely to “like” a picture in Facebook if many of her friends have done so). Flickr’s interestingness attribute is another example of a community-driven measure of appeal based on user-judged content and community reinforcement. Flickr honors ideas and imagination in addition to visual appeal within its interestingness measure.

A user study to determine factors that would prevent people from including a picture in their albums was reported in [76]. Factors such as “not an interesting subject,” “a duplicate picture,” “occlusion,” or “unpleasant expression” were found to dominate the list. Attributing multidimensional image value indexes (IVIs) to pictures based on their technical and aesthetic qualities and social relevance has been proposed in [60]. While technical and aesthetic IVIs are driven by learned models based on low-level image information, an intuitive social IVI methodology can be adherence to social rules learned jointly from users’ personal collections and social structure. An example could be to give higher weights to immediate family members than cousins, friends, and neighbors in judging a picture’s worth [60].

DISCUSSION

While a personal or situational appeal or value would be of greater interest to a nonspecialist user, generic models for appeal may be even more short-lived than for aesthetics. To make an impact, the problems within this category must be carefully tailored toward learning personal or situational preferences. From an algorithmic perspective, total dependence on visual characteristics, for modeling and predicting consumer appeal, is a poor choice and it is desirable to employ image metadata such as tags, geographical information, time, and date. Inferring relationships between people based on the faces and their relative geometric arrangements in photos could also be a very useful exercise [37].

AESTHETICS AND EMOTIONS IN ARTWORK CHARACTERIZATION

Artistic use of paint and brush can evoke a myriad of emotions among people. These are tools that artists employ to convey their ideas and feelings visually, semantically, or symbolically. Thus they form an important part of the study of aesthetics and emotions as a whole. Painting styles and brushstrokes are best understood and explained by art connoisseurs. However, research in the last decade has shown that models built using low-level visual features can be useful aids to characterize genres and painting styles or for retrieval from large digitized art galleries [13], [14], [29], [52], [53], [75]. In an effort to encourage computational efforts to analyze artwork, the Van Gogh and Kröller-Müller Museums in The Netherlands have made 101 high-resolution grayscale scans of paintings available to several research groups [45].

Brushstrokes provide reliable modeling information for certain types of paintings that do not have colors. In [58], mixtures of stochastic models have been used to model an artist's signature brushstrokes and painting styles. The research provides a useful methodology for art historians who study connections among artists or periods in the history of art. Another important formulation of this characterization problem has been discussed in [12]. The work constructs an artists' graph wherein the edges between two nodes are representative of some measure of collective similarities between paintings of the two artists. It is shown that the connections uncovered with the graph are coherent with intuitive judgment and statements of art specialists about the paintings and artists in question. In addition to these, influences of artists on one another are also captured and represented in the graph. While a connoisseur's view of art may be valuable, another valuable problem to the commercial art community is to model and predict a common-man's perception and appreciation of art. This has been attempted in a more recent work [57], which attempts to determine aesthetic quality of paintings based on ground truth obtained from common people as opposed to art connoisseurs.

While it is natural for humans to interpret facial expressions, computer vision algorithms have proven of late largely worthy of the same. An interesting application of facial expression recognition technology has been shown to be the decoding of the expression of portraits such as the *Mona Lisa* to get an insight into the artists' minds [117]. Understanding the emotions that paintings arouse in humans is yet another aspect of this research. A method that categorizes emotions in art based on ground truth from psychological studies has been described in [102]. The authors of the cited work present a cross-domain application where training is performed using a well-known image data set in psychology while the approach is demonstrated on certain art masterpieces.

DISCUSSION

Problems discussed within this category range from learning nuances of brushstrokes to emotions that artworks arouse in humans and even emotions depicted in the artworks themselves

(especially portraits like the *Mona Lisa*). This is a challenging area and the research is expected to be helpful to curators of art as well as to commercial art vendors. However, contribution here would, in most scenarios, benefit from direct input of art experts or artists themselves. As most of the paintings that are available in museums today were done before the 20th century, obtaining first-hand input from artists is impossible. However, such research aims to build healthy collaborations between the art and computer science research communities, some of which are already evident today [45].

AESTHETICS, EMOTIONS, AND ATTRACTIVENESS

Another manifestation of emotional response is attraction among human beings especially to members of the opposite sex. While the psychology of attraction may be multidimensional, an important aspect of attraction is the perception of a human face as beautiful. Attraction is an attribute resulting from a complex mixture of emotions and perceived aesthetics, usually inexplicable, and therefore understanding and assessment of attractiveness is a research problem in its own right. Understanding beauty has been an important discipline in experimental psychology [93]. Traditionally, beauty was synonymous with perfection and hence symmetric or perfectly formed faces were considered attractive. In later years, psychologists conducted studies to indicate that subtle asymmetry in faces is perceived as beautiful [77], [86], [103]. Therefore, it seems that computer vision research on asymmetry in faces, such as [59], can be integrated with psychological theories to computationally understand the dynamics of attractiveness. Another perspective is the theory that facial expression can affect the degree of attractiveness of a face [24]. The cited work uses advanced magnetic resonance imaging (MRI) techniques to study the neural response of the human brain to a smile. The current availability of Web resources has been leveraged to formulate judging facial attractiveness as a machine learning problem [23].

DISCUSSION

Research in this area is tied to work in face and facial expression recognition. There are controversial aspects of this research in that it tries to prototype attraction or beauty by visual features. While it is approached here purely from a research perspective, the overtones of the research may not be well accepted by the community at large. Beauty and attraction are personal things and many people would dislike it to be rated on a scale. It should also be noted that so-called beauty contests also assess the complete personality of participants and do not judge merely by visual aspects.

AESTHETICS, EMOTIONS, AND IMAGE RETRIEVAL

While image retrieval largely involves generic semantics modeling, certain interesting offshoots that involve feedback, personalization, and emotions in image retrieval have also been studied for several years [94]. Introducing such a personal or human touch into retrieval is expected to produce more rewards. Problem formulation undergoes little change and the

goal is still to retrieve the most relevant pictures given a keyword or a query image. However, human factors such as those mentioned above provide a useful way to rerank or search among equals for matches closer to the heart of a user. In [10], an image filtering system that uses the Kansei user model has been described. The Kansei user model has its roots in Kansei engineering, which deals with translating feelings and impressions into product parameters. From an image modeling perspective, the Kansei methodology should encompass methods that associate low-level image features with human feelings and impressions. Another work [30] attempts to model the target image within the mind of a user with respect to a face retrieval task. In the cited work, a relevance feedback-based approach is used to learn a distribution over the image database that represents the mental image of the user, and to use this distribution for retrieval.

DISCUSSION

Image retrieval is itself a vast research area. Of late there is emphasis on human centered multimedia information processing, which also touches aspects of retrieval. However, such research is not easily evaluable or verifiable as again the level of subjectivity is very high. While it still remains an important question as to how much commercial benefit a totally personalized human-centered image retrieval system would yield over a generic semantics understanding retrieval system, research in this direction is definitely valuable from an academic standpoint. In particular, the tradeoff between personalization and speed needs to be explored.

COMPUTATIONAL FRAMEWORKS

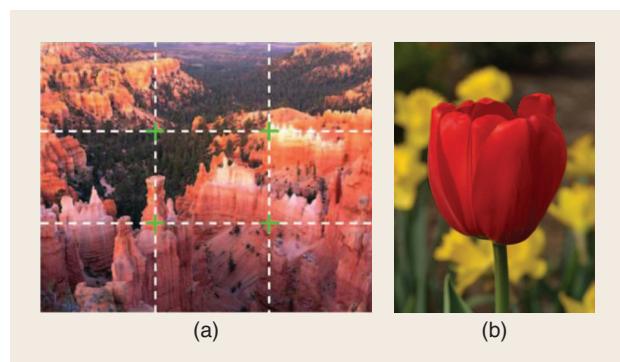
With the problem descriptions in place, a framework to describe the distinct procedures taken to address problems in this domain is impending. From a computational perspective, we need to consider steps that are necessary to obtain a prediction (some function of the aesthetics or emotional response) from an input image. We divide this discussion into two distinct sections, “Feature Representations” and “Machine Learning,” and elucidate how researchers have approached each of these computational aspects with respect to the current field. However,

before moving forward, it is important to understand and appreciate certain inherent gaps when any image understanding problem is addressed in a computational way. Smeulders et al. introduced the term *semantic gap* in their pioneering survey of image retrieval to summarize the technical limitations of image understanding [81]. In an analogous fashion, the technical challenge in automatic inference of aesthetics is defined in [22] as the aesthetics gap, as follows: The aesthetics gap is the lack of coincidence between the information that one can extract from low-level visual data (i.e., pixels in digital images) and the aesthetics response or interpretation of emotions that the visual data may arouse in a particular user in a given situation.

FEATURES AND REPRESENTATION

In the last decade and a half, there have been significant contributions to the field of feature extraction and image representation for semantics and image understanding [20]. Feature extraction and image representation are prerequisites to any image understanding task, and aesthetics or emotional inference are no exceptions (and in some sense more critical). Aesthetics and emotional values of images have bearings on their semantics and so it is not surprising that feature extraction methods are borrowed or inspired from the existing literature. There are psychological studies that show that aesthetic response to a picture may depend upon several dimensions such as composition, colorfulness, spatial organization, emphasis, motion, depth, or presence of humans [5], [34], [71]. Conceiving meaningful visual properties that may have correlation with perceived aesthetics or an emotion is itself a challenging problem. In the literature, we notice a spectrum from very generic color, texture, and shape features to specifically designed feature descriptors that are expected to capture the perceptual properties that contribute to the aesthetic or emotional value of a picture or artwork. We do not intend to provide an exhaustive list of feature descriptors here but rather discuss significant feature usage patterns.

Photographers generally follow certain principles that can distinguish professional shots from amateur ones. A few such principles are the rule of thirds, use of complementary colors, and close-up shots with high dynamic ranges. The rule of thirds is a popular one in photography. It specifies that the main element or the center of interest in a photograph should lie at one of the four intersections (Figure 4). In [19], the degree of adherence to this rule is measured as the average hue, saturation, and intensities within the inner third region of a photograph. It has also been noted that pictures with simplistic composition and a well-focused center of interest are more pleasing than pictures with many different objects. Professional photographers often reduce the depth of field (DOF) to shoot single objects by using larger aperture settings, macro lenses, or telephoto lenses. DOF is the range of distance from a camera that is acceptably sharp in a photograph (Figure 4). In [19], wavelets have been used to detect a picture with a low DOF. However, low DOF has a positive aesthetic appeal only in an appropriate context and may not always be desirable (e.g., in photography, landscapes with



[FIG4] (a) The rule of thirds in photography. (b) A low depth-of-field picture.

narrow DOF are not considered pleasing; instead, photographers prefer to have the foreground, middle ground, and background all in focus).

A mix of global and local features has been used in [57] to model the aesthetics problem for paintings. Feature selection is based on the belief that people use a top-down approach to appreciate art. A more holistic impression is first gathered, followed by perusal of the details. Prominent factors that determine the choice of features include measuring blur (which is seen as an important artistic effect) and presence and distribution of edges, because edges are used by artists for emphasis. The perceptual qualities that differentiate professional pictures from snapshots based on input from professional and amateur photographers are identified in [47]. It is found that professional shots are distinguished by i) a clear distinction between subject and background brought about by choice of complementary colors, higher contrast between subject and background, or a small DOF, and ii) a surrealism created by the proper choice of camera parameters and appropriate lighting conditions. Conversely, a largely blurred or a low-contrast picture is likely to be a snapshot by an amateur.

While low-level color and texture features capture useful information, modeling spatial characteristics of pixels or regions and spatial relationships among regions in images has been shown to be very helpful. A computational visual attention model using a face-sensitive saliency map is proposed in [85]. A rate of focused attention measure (using the saliency map and the main subject of the image) is proposed as an indicator of aesthetics. The method employs a subject mask generated using several hundreds of manually annotated photos for computation of attention. Yang et al. propose an interesting pseudogravitational field-based visual attention model in [101] where each pixel is assigned a mass based on its luma and chroma values (YCbCr space) and pixels exert a gravity-like mutual force. An iterative algorithm that employs this gravitation model computes fixation points.

Some recent papers focus on enhancement of images or suggestion of ideal composition based on aesthetically learned rules [11], [17]. Two distinct recomposition techniques based on key aesthetic principles (rule of thirds and golden ratio) have been proposed in [11]. The algorithm performs segmentation of single subject images into “sky,” “support,” and “foreground” regions. Two key aesthetically relevant segment-based features are introduced in this work; the first computes the position of the visual attention center with respect to focal stress points in the image (rule of thirds), while the second feature measures the ratio of weights of support and sky regions (expected to be close to golden ratio). Using learned classifiers and an inpainting algorithm, users are suggested optimal positioning of subjects within the image frame, or are prompted to readjust sky and support regions in natural images. Yet another interesting work [17] models local and far contexts from aesthetically pleasing pictures to determine rules that are later applied to suggest good composition to new photographers. According to the authors, while local context represents visual continuity, far context models the

arrangement of objects/regions as desirable by expert photographers. Images are segmented using a graph-based algorithm into regions and a visual vocabulary is constructed. Contextual modeling involves learning a spatial Gaussian mixture model for pairwise visual words.

While there exists some concrete rationalization for feature design with respect to the aesthetics inference problem, designing features that capture emotions is still a challenge. In [102], the emotion categorization problem in art is considered using simplistic visual features. In the cited work, the authors, however, divert from the common codebook approach to a methodology where similarity to all vocabulary elements is preserved for modeling. Weibull distribution is used to model color invariant edges and Gabor filters are used to measure the surface texture. In [12], low-level local visual features including scale-invariant feature transform and color histograms are extracted and a Fisher Kernel-based image similarity is used to construct a graph of artists to discover mutual and collective artistic influence. Associating low-level image features with human feelings and impressions can also be achieved by using ideas from Kansei engineering [10]. The authors of the cited work use sets of neural networks, which try to learn mappings between low-level image features and high-level impression words.

Concepts from psychological studies and art theory are used to extract image features for emotion recognition in images and art in [65]. Among other features, [65] adopts the standardized pleasure-arousal-dominance transform color space, composition features such as low-DOF indicators and rule of thirds (which have been found to be useful for aesthetics), and proportion of skin pixels in images. In [73], eye-gaze analysis yields an affective model for objects or concepts in images. More specifically, eye fixation and movement patterns learned from labeled images are used to localize affective regions in unlabeled images. Affective responses in the form of facial expressions are explored in [3] to understand and predict topical relevance. The work models neurological signals and facial expressions of users looking at images as implicit relevance feedback. To classify emotions, [3] employs a 3-D wire-frame model of faces and tracks presence and degrees of changes in different facial regions. Similarly, [92] also employs face tracking to extract facial motion features for emotion classification.

Finally, psychological theories of perception of beauty (discussed previously) also aid researchers who design features for facial attractiveness modeling using a mix of facial geometry features [23], [28] as well as nongeometric ones (such as hair color and skin smoothness) [28].

MACHINE LEARNING

Learning lies at the heart of every computational inference problem that we consider here. The choice of the learning strategy, however, depends upon the nature of the problem and the task to be achieved. Here, we describe the following important dualities in learning paradigms and lay out scenarios, within our discussion scope, which should guide the choice of an appropriate strategy.

SUPERVISED VERSUS UNSUPERVISED LEARNING PARADIGMS

While supervised learning methods are used in the presence of ground truth to learn classification patterns among data, unsupervised methods can learn patterns among data in a more impromptu fashion. The major distinction that guides the choice here is the availability of ground truth. In the absence of ground truth, for unsupervised methods, feature similarity between data points is the driving factor for pattern discovery. Supervised learning has been used for aesthetics inference using support vector machines (SVMs) and classification and regression trees (CART) schemes in [19], [57], and [60], and for emotion inference using SVMs in [102]. An example of unsupervised learning within our scope is the construction of the painters' graph in [12], which is then used to infer connections between painter styles and genres. Elements of unsupervised learning in the form of i) K-means clustering for visual vocabulary generation, ii) graph-based region segmentation, and iii) image clustering to form topical groups, are found in [17]. While supervised learning methods may form the cornerstone for classification in most problems, unsupervised methods have often been found to successfully achieve intermediate tasks (such as those discussed above).

GENERATIVE VERSUS DISCRIMINATIVE LEARNING PARADIGMS

The generative learning philosophy assumes that some underlying statistical process can generate the observed data and the goal is to learn the process. On the other hand, discriminative learning operates on no such assumption and learns class-specific rules or mathematical space structures given the data and class information. Typically, discriminative learning methods are supervised in nature and divide the data space using hard boundaries, whereas the division in generative learning methods is softer and probabilistic. Generative learning methods are usually computationally lightweight in nature but over the years, discriminative learning has in general proven to be more widely and efficiently used. Not surprisingly, most instances of learning for aesthetics and emotions inference use discriminative strategies such as SVMs and CART [19], [57], [60], [102]. However, there are several examples of generative learning in literature such as the use of Bayesian networks in [63] to select the most appealing image in an event. The Naïve Bayes approach is also used in [65] for emotion modeling compared to sophisticated SVMs and trees, which they find to be best in terms of performance and speed. A recent work, [92] employs a Bayesian network classifier to classify facial expressions into emotional categories such as angry, disgusted, or happy. The spatial context between pairs of visual words is modeled using a Gaussian mixture generative modeling approach in [17].

TWO-CLASS VERSUS MULTICLASS CLASSIFICATION PARADIGMS

A two-class problem is the most easily formulated and widely studied classification problem in learning literature. However, in a realistic scenario, there could be more than two classes to which data

can potentially belong. CART-based classification assumes the presence of multiple classes. When using SVMs as classifiers, most multiclass cases can be formulated as some extensions of the two-class problem. Among various formulations, the prominent ones include "one-versus-all" and "one-versus-one" classification. Aesthetics inference in [19] and [57] is treated as a two-class classification problem where the two classes are "high score" and "low score," respectively. This can be considered a plausible choice because intuitively it is difficult to distinguish between small variations in user ratings. It has also been found that if the score-gap between the two classes is made wider, the classification performance improves. Emotion recognition has been formulated as a multiclass classification problem in [102] and [65]. In general, there are always more than two classes (aesthetics or emotions) present in data. However, for aesthetics ratings (one to seven), the classes that correspond to say score three and four may not be quite independent. To accommodate multiple aesthetics ratings, a more appropriate formulation is regression (as discussed below). Moreover, a multiclass classification problem in general calls for employing contextual relationships between class semantics to boost classification performance. Relationships between emotions are yet to be explored under this framework.

CLASSIFICATION VERSUS REGRESSION PARADIGMS

In a classification paradigm, an important assumption is that the data can only belong to one or a finite number of classes or categories, and the goal is to discover class boundaries in the data space. An alternate paradigm, regression, allows the data to be associated with real numbers (such as aesthetics ratings) with the aim to learn some form of a mathematical function that can efficiently associate the data with the real number space. The bulk of the formulations in the emotions and aesthetics inference sphere employ classification [19], [57], [102]. This is reasonable for emotion inference because of the finiteness and clear characterization of the human emotional space. Conversely, aesthetics is a more abstract quality and quantifying it requires a relatively larger numeric scale. However, this naturally results in greater variations in user ratings, making the learning task prone to noise. A regression formulation of the aesthetics inference problem has been studied in [19] where the regression function attempts to learn raw user ratings (as opposed to classes such as high-score or low-score). A support vector regression framework to learn and predict aesthetic quality of single subject images and suggest ideas for recomposition has been proposed in [11]. Human attractiveness prediction has been modeled as a manifold kernel regression problem in [23].

APPLICATIONS TO REAL-WORLD SYSTEMS

Thus far, we have described the research problems in aesthetics and emotions inference and discussed approaches to them. In this section, we focus on how solutions to the problems can have real-world impact. Roughly speaking, the impact can be thought to be in the form of improving the user experience of existing technological systems, and more broadly, playing a role to improve the present-day quality of life. Here, we consider three broad areas of real-world impact, i.e., photographic systems, image search systems, and indirect areas of application.

PHOTOGRAPHIC SYSTEMS

Expertise in photographic judgment of quality is arguably a skill acquired over time and through a large amount of exposure. However, every other person possesses a digital camera these days, taking hundreds of new photographs on every occasion. While an expert photographer can probably sort out the good-quality photographs from the bad ones during a postevent analysis, the photographs chosen can only be subsets of those taken. This brings out the inherent problem of postprocessing a photo collection to obtain high-quality moments. We envision a future where consumer cameras are equipped with an automated personal assistant that can help capture moments at the instant they occur, so that only the highest quality photos are taken and stored, and postanalysis becomes unnecessary. While this may seem unlikely, there is reason to believe that we may achieve something like this in the near future. In both [19] and [47], the software side of visual aesthetics judgment has been explored extensively, and it has been argued that these aesthetics judgment models can be converted into hardware systems and embedded into consumer cameras. There could be at least three different kinds of embedded aesthetics judgment modules onboard cameras:

POSTPHOTOGRAPHY FILTER

After a few shots are taken, one could press a button to activate the aesthetics module, which then filters out poorly taken photos or retains the best few shots.

Advantage

The quality judgment can be done on a relative scale with the remainder of the photos in storage, thereby allowing one to choose the best few or the worst few photographs. A software-based analogue of this type of a filter was proposed in [21].

Drawback

The main drawback of such a module is that it cannot save effort (and storage) spent taking poor-quality photographs by warning the user at the time they are taken. Second, this functionality onboard a camera may not be as attractive as the software could instead be part of the photo upload tool on a computer, thereby saving the cost of embedding while retaining the same functionality.

REAL-TIME FILTER

A real-time filter onboard a camera, when activated, monitors potential shots that are taken as the camera is actively used to find potential shots. As a result, an "aesthetics meter" could reflect the expected aesthetic value as it is estimated, and hence allow for real-time adjustments to the camera pose and settings, so a high-quality shot more likely results at the onset.

Advantage

Assuming that the filter is reliable, the ability to track photo quality and hence to take the best shot at the time it is taken can save much effort. Unlike a postphotography filter, there is

no risk of not having taken the best possible shot one could have in a particular event or scenario. When camera storage is limited, this can be particularly useful for adding new photographs in a calculated manner.

Drawbacks

One of the main drawbacks of a real-time system could be that the need to generate a low-latency aesthetics score may force a lightweight algorithm, thereby compromising on the quality of the metric. It also prevents a relative comparison of photographs taken at various scenes so as to retain a sampling of high-quality photos or eliminate low-quality ones, as proposed in [21].

While this may seem an interesting application area of aesthetic quality scoring, it is inherently controversial. For one, a computational feedback on a continuous basis for something that is highly subjective can arguably take away the pleasure derived from traditional photography. Also, given that such systems are unlikely to reach high levels of accuracy in the near future, the errors they make can have serious consequences if followed strictly. This feature thus appears more appropriate for the novice photographer than the expert, as well as an aide to the spectrum of people between the two categories. Experts may, however, explore this technology differently, to arrive at new compositions by denying, disturbing, or challenging the feedback from such cameras.

Nevertheless, onboard aesthetic quality meters are a speculative, novel, and largely unexplored territory. User experience can be truly assessed only when such systems become available in the consumer market. A recent effort in this direction is the Nadia camera that uses an offline aesthetics prediction engine (ACQUINE) to offer a real-time aesthetics score [114].

REAL-TIME FEEDBACK

A real-time filter onboard a camera, when activated, monitors a shot just taken, although the camera is not actively used to find shots. As a result, an aesthetics meter could reflect the aesthetic value almost instantly, and hence allow for real-time retake of the picture through user adjustments to the camera pose and settings, so as to take a high-quality shot on the spot. This is an in-between option that has already been implemented in cameras, e.g., in the form of a blurry picture warning (red-blurry, yellow-questionable, green-no blur). Research by Barry et al. [7], [8], [80] in the domain of providing intelligent real-time commonsense feedback to videographers is a good example of work in this direction. The cited work explores how commonsense knowledge about events and expert event capture can be prompted to an amateur videographer in the form of suggestions that enhance video taking.

IMAGE SEARCH AND ORGANIZATION SYSTEMS

Image search systems have historically focused on search relevance. In particular, content-based image retrieval systems have used precision-recall metrics as the de facto standards for comparing algorithms [20]. As with the Web over the years, there has

been an explosion of digital content to be indexed. The effect of that has been that for most common search queries, there are a large number of relevant results from which to choose. With images, after a point the relevance ranking functions are likely arbitrary. Given that a large number of images are known to be nearly equivalent to each other in semantics, one way to rank them is by their aesthetic quality. This particular area of image retrieval has recently begun to generate interest. However, along with relevance and quality, there is another metric that has been mentioned repeatedly in the literature in conjunction with image search diversity. Because it is well known that users only look for search results in the top few ranked pages, if the different kinds of images that pertain to the query do not appear within the top few pages, users may be disappointed. There are therefore at least three types of metrics that can play a role in image search result ranking: relevance, aesthetic quality, and diversity. Given that relevance has been addressed and we are only dealing with images that are of interest to the user, the question is how we can promote some images up the ranks of some others. Some reranking can be done based on simple factors such as size and shape of the image. Then, the trickier part is the balance of aesthetic quality with diversity. The reason that diversity plays an important role in the context of aesthetics-based ranking is that it may occur that certain types of images are inherently more aesthetically pleasing than others. As a result, the image-ranking function may prevent a diverse selection of images from appearing near the top of the array of selections. To enforce diversity while still focusing on relevance and quality, a simplified algorithm can be as follows:

- 1) Let X be the set of all images in corpus.
- 2) For query Q , generate subset X' of X containing relevant images only.
- 3) Based on a diversity metric, cluster X' into diverse sets of images.
- 4) Within each diverse cluster, rank images by their aesthetic quality.
- 5) Show top-ranked images from each diverse cluster, ranked overall by a combination of the three metrics.

While this is a fairly generic algorithm, the details should be completed by conducting user studies and determining the right mix of these three metrics that lead to good user satisfaction, or alternatively, allowing users to specify their personal preferences.

DATA RESOURCES

DATA FROM CONTROLLED STUDIES

Methods for experimental investigation of aesthetic perception and preferences and associated emotional experience vary from traditional collection of verbal judgments along aesthetic dimensions, to multidimensional scaling of aesthetic value and other related attributes, to measuring behavioral, psychophysiological, and neurophysiological responses to art pieces and images in controlled and free viewing conditions. The arsenal of measured response is vast, a few instances being reaction time, various electrophysiological responses that capture activity of the central and autonomic nervous systems,

such as an electroencephalogram, electrooculogram, heart rhythm, pupillary reactions, and more recently, neural activity in various brain areas obtained using functional MRI [50], [24]. Recording eye movements is also a valuable technique that helps detect where the viewers are looking when evaluating aesthetic attributes of art compositions [68].

Certain efforts have resulted in the creation of a specialized database for emotion studies known as the International Affective Picture Systems database [55]. The collection contains a diverse set of pictures that depict animals, people, activities, and nature, and has been categorized mainly in valences (positive, negative, no emotions) along various emotional dimensions [102].

DATA FROM COMMUNITY-CONTRIBUTED RESOURCES

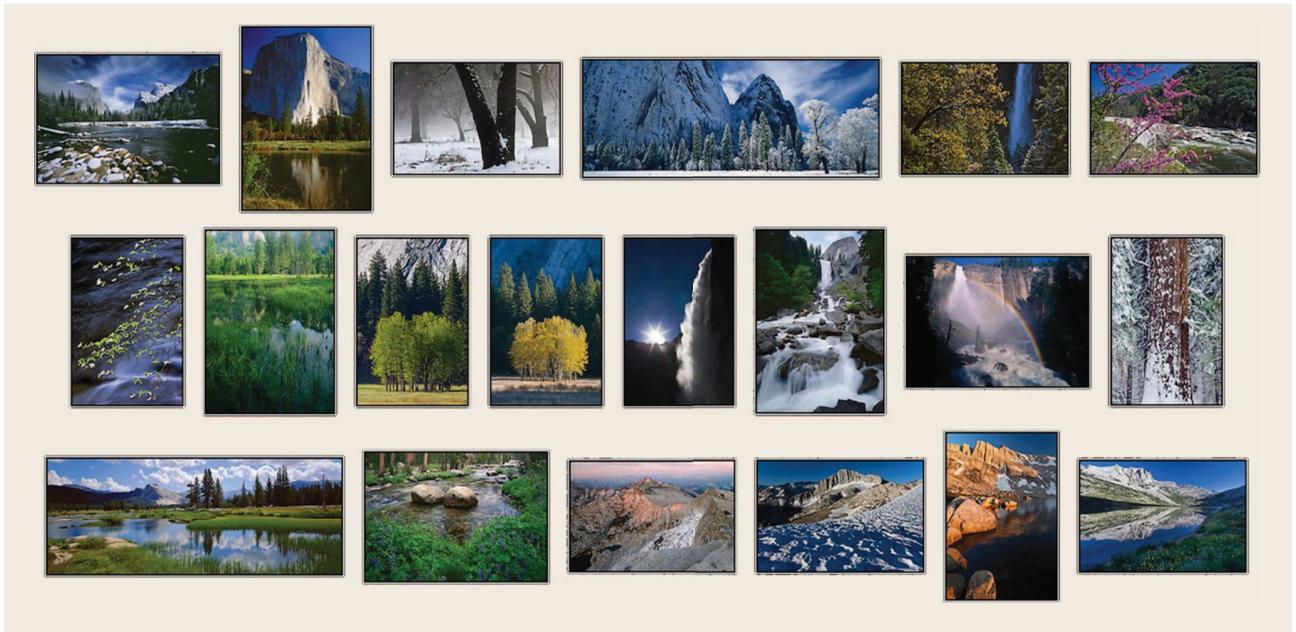
Obtaining controlled experimental data is expensive in time and cost. At the same time, converting user response (captured as described above) to categorical or numerical aesthetics or emotional parameters is another challenge. One should also note that controlled studies are not scalable in nature and can only yield limited human response in a given time. Researchers increasingly turn to the Web, a potentially boundless resource for information. In the last few years, a growing phenomenon called crowd sourcing has hit the Web. By definition, crowd sourcing is the process by which Web users contribute collectively to useful information on the Web [44]. Several Web photo resources take advantage of these contributions to make their content more visible, searchable, and open to public discussions and feedback. Tapping such resources has proven useful for research in our discussion domain. Here we briefly describe some Web-based data resources.

- *Flickr* [113] is one of the largest online photo-sharing sites in the world. Besides being a platform for photography, tagging, and blogging, Flickr captures contemporary community interest in the form of an interestingness feature. According to Flickr, the interestingness of a picture is dynamic and depends on a plurality of criteria including its photographer, who marks it as a favorite, comments, and tags given by the community.

- *Photo.Net* [115] is a platform for photography enthusiasts to share and have their pictures peer rated on a one to seven scale of aesthetics. The photography community also provides discussion forums, reviews on photos and photography products, and galleries for members and casual surfers.

- *DPChallenge* [111] allows users to participate and contest in theme-based photography on diverse themes such as life and death, portraits, animals, geology, street photography. Peer rating on overall quality, on a one to ten scale, determines the contest winners.

- *Terragalleria* [116] showcases travel photography of Quang-Tuan Luong (a scientist and a photographer), and is one of the finest resources for U.S. national park photography on the Web (Figure 5). All photographs here have been taken by one person (unlike Photo.Net), but multiple users have rated them on overall quality on a one to ten scale.



[FIG5] Pictures of Yosemite National Park from Terragalleria.com (used with permission).

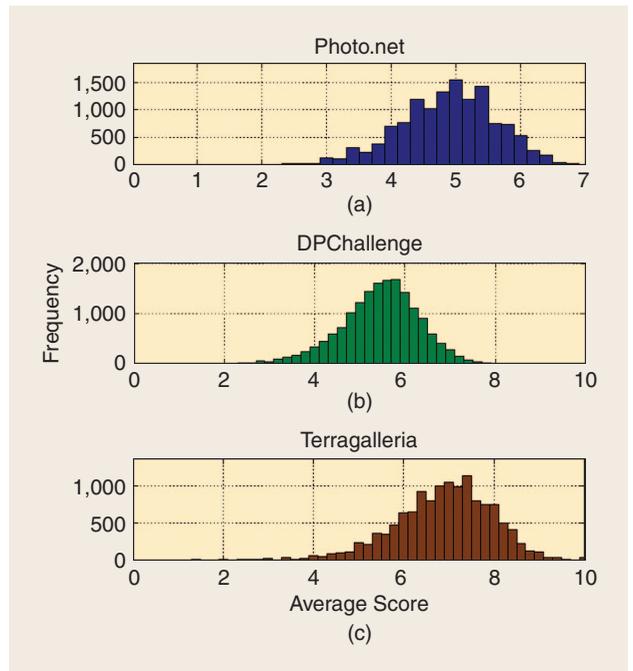
■ ALIPR [110] is a Web-based image search and tagging system that also allows users to rate photographs along ten different emotional categories such as surprising, amusing, pleasing, exciting, and adorable.

DATA ANALYSIS

FEATURE PLOTS OF AESTHETICS RATINGS

We performed a preliminary analysis of the above data sources to compare and contrast the different rating patterns. A collection of images (14,839 images from Photo.net, 16,509 images from DPChallenge, 14,449 images from Terragalleria, and 13,010 emotion-tagged images from ALIPR) was formed, drawing at random, to create real-world data sets (to be available at <http://riemann.ist.psu.edu/>). These can be used to compare competing algorithms in the future. Here we present plots of features of the data sets, in particular the nature of user ratings received in each case (not necessarily comparable across the data sets).

We first describe the nature of the plots. In the following section, we conduct a thorough analysis of each figure, breaking it up for each data source/quality score received by each photo. Figure 6 shows the distribution of mean aesthetics. Figure 7 shows the distribution of the number of ratings each photo received. In Figure 8, the number of ratings per photo is plotted against the average score received by it, in an attempt to visualize possible correlation between the number of ratings and the average ratings each photo received. In Figure 9, we plot the distribution of the fraction of ratings received by each photo within ± 0.5 of its own average. In other words, we examine every score received by a photo, find the average, count the number of ratings that are within ± 0.5



[FIG6] Parts (a)–(c) show distributions of average aesthetics scores from three different data collections.

of this average, and take the ratio of this count and the total number of ratings this photo received. This is the ratio whose distribution we plot. Each of the aforementioned figures comprises this analysis separately for each collection (Photo.net, Terragalleria, and DPChallenge). Finally, in Figure 10, we plot the distribution of emotions votes in the data set sampled from ALIPR. In the following section, we will analyze each of

these plots separately and share with readers the insights drawn from them.

ANALYSIS OF FEATURE PLOTS

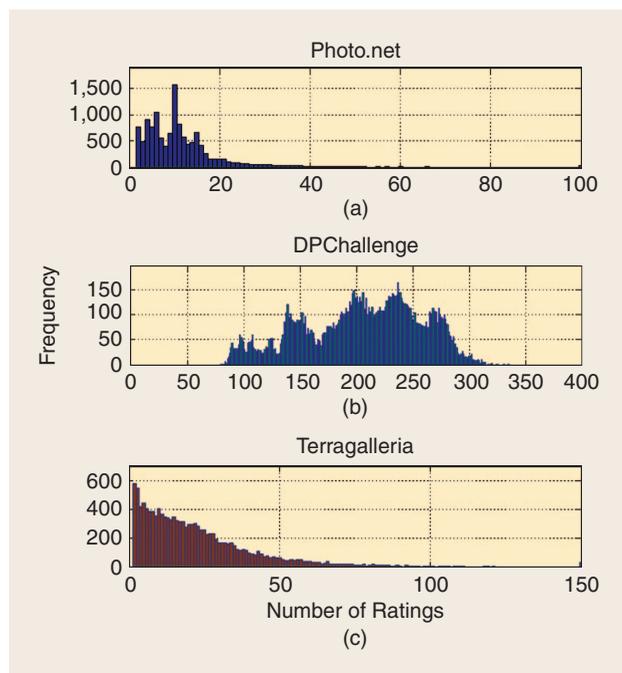
When we look closely at each of the plots in Figures 6–10, we obtain insights about the nature of human ratings of aesthetics. Broadly speaking, we note that this analysis pertains to the overall social phenomenon of peer rating of photographs rather than the true perception of photographic aesthetic quality by individuals. In Photo.net, for example, users (at least at the time of data collection) could see who rated their photographs. This naturally makes the rating process a social rather a true scientifically unbiased test or process. Another side effect of this is

that the photos that people upload for others to rate are generally not drawn at random from a person’s broad picture collection. Rather, it is more likely that they select to share what they consider their best taken shots. This introduces another kind of bias. Models and systems trained on this data therefore learn how people rate each other’s photos in a largely nonblind social setting, and only learn this for a subset of the images that users consider worthy of being posted publicly. Bearing this in mind helps to explain the inherent bias found in the distributions. Conversely, the bias corroborates the assumption that collection of aesthetics rating in public social forums is primarily a social experiment rather than a principled scientific one.

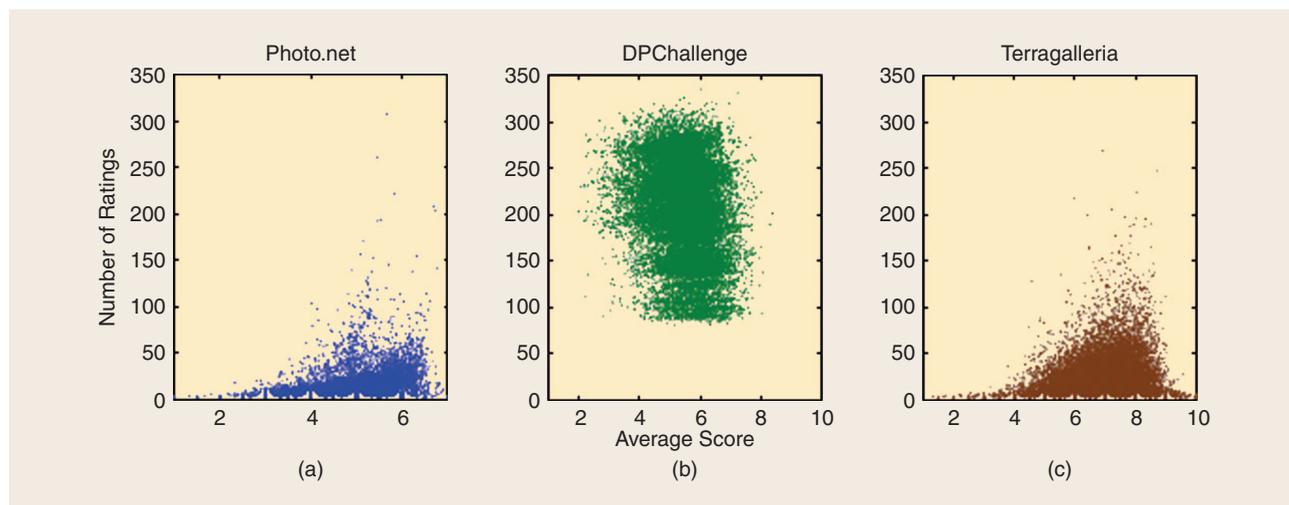
In Figure 6, we see that for each data set, the peak of the average score distribution lies to the right of the mean position in the rating scale. For example, the peak for Photo.net is approximately five, which is a full point above the midpoint four. There are two possible explanations for this phenomenon:

- Users tend to post only those pictures that they consider to be their best shots.
- Because public photo rating is a social process, peers tend to be lenient or generous by inflating the scores that they assign to others’ photos, as a means of encouragement and also particularly when the Web site reveals the rater’s identity.

Another observation we make from Figure 6 is that the distribution is smoother for DPChallenge than for the other two. This may simply be because this data set has the largest sample size. In Figure 7, we consider the distribution of the number of ratings each photo received. This graph looks dramatically different for each source. This feature almost entirely reflects on the social nature of public ratings rather than anything intrinsic to photographic aesthetics. The most well-balanced distribution is found in DPChallenge, in part because of the incentive structure (it is a time-critical, peer-rated competitive platform). The distribution almost resembles a mixture of Gaussians with means at well-spaced locations. It is unclear to the authors with which social phenomenon on DPChallenge.com these peaks might be associated. Photos on Photo.net are



[FIG7] Parts (a)–(c) show distributions of number of ratings from three different data collections.



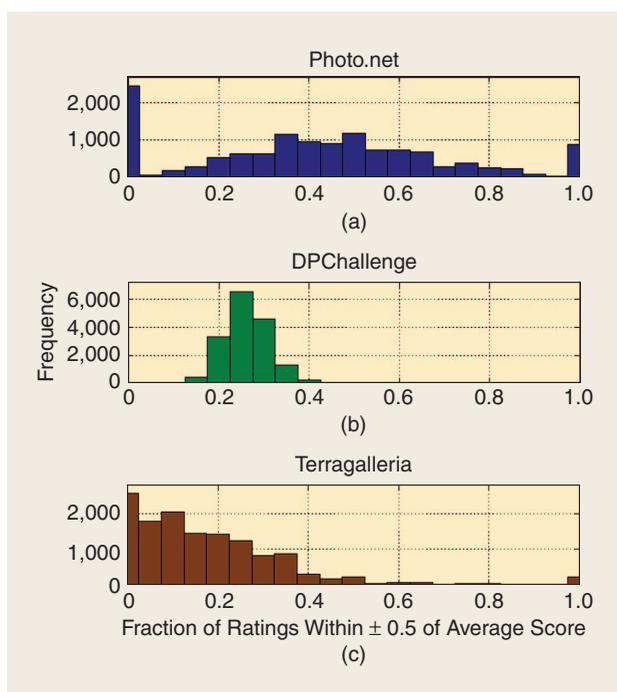
[FIG8] Parts (a)–(c) show the correlation plots of (average score, number of ratings) pairs.

much rarer, mainly because the process is noncompetitive, voluntary, and the system of soliciting ratings is not designed to attract many ratings per photo. The distribution looks heavy tailed in the case of Terragalleria, which much more resembles typical rating distribution plots.

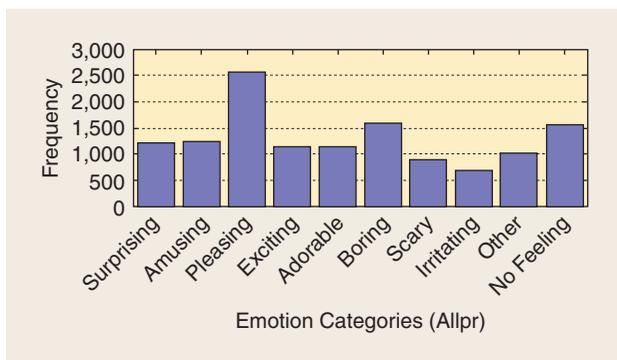
The purpose of the plots in Figure 8 is to determine if there exists a correlation between the number of ratings a photo receives and the average of those ratings. The plots for Photo.net as well as Terragalleria most clearly demonstrate what can be anticipated about social peer-rating systems: people rate inherently positively, and they tend to highly rate photos that they like, and not rate at all those they consider to be poor. This phenomenon is not peculiar to photo-rating systems or even social systems: we also observe this clearly in movie rating systems found on Web sites such as IMDB. Associated with the issue that people tend to explicitly rate mainly things they like is the fact that the Web sites also tend to surface highly rated entities to newer audiences (through top K lists and recommendations). Together, these two forces help generate much data on good-quality entities while other candidates are left with sparse amounts of feedback and rating. Conversely, DPChallenge, because it is a competitive site, attempts to fairly gather feedback from all candidate photos. Therefore, we see a less biased distribution of its scores, making it unclear whether the correlation is at all significant or not.

In Figure 9, we plot the distribution of the fraction of ratings received by each photo within ± 0.5 of its own average. What we expect to see is whether or not most ratings are closer to the average score. In other words, do most raters roughly agree with each other for a given photo, or is the variance per photo high for most photos? The observation for Photo.net is that there is a wide and healthy distribution of the fraction of rater agreement, and then there are the boundary conditions. A small but significant fraction of the photos had everyone essentially give the photo the same rating ± 0.5 (this corresponds to $x = 1$ in the plot). These photos have high consensus or rater agreement. However, three times larger is the fraction of photos where nearly no one has given a rating close to the average (this corresponds to $x = 0$ in the plot). This occurs primarily when there are two groups of raters: one group that likes the photo and another group that does not. This way, the average lies somewhere between the sets of scores given by the two camps of raters. The distribution looks quite different for DPChallenge: roughly one third of the ratings tend to lie close to the average value, while the rest of the ratings lie further apart on either side of average. For Terragalleria, users tend to be less in agreement with each other on ratings. Nearly all of the raters are in agreement on only a small fraction of the photos (corresponding to $x = 1$ in the plot).

Note that the graphs in Figure 9 are particularly unfit for an apples-to-apples comparison: an absolute difference of 0.5 implies different things for the different Web sites, especially since the score ranges are different. Furthermore, DPChallenge receives so many ratings per photo that it is improbable that all raters would agree on the same score (hence $y = 0$ at $x = 1$ in



[FIG9] Parts (a)–(c) show the distributions of the level of consensus among ratings.



[FIG10] Distribution of emotion votes given to images (ALIPR).

that graph). Finally, in Figure 10, we observe that the dominant emotion expressed by Web users while viewing pictures is “pleasing,” followed by “boring,” and “no feeling.” Conversely, “irritating” and “scary” are relatively rare responses. The reason for this may well be what emotions people find easy to attribute to the process of looking at a picture. On the Web, we are accustomed to expressing ourselves on like-dislike scales of various kinds. Hence, it is convenient to refer to what one likes as “pleasing” and what one does not like as “boring.”

FUTURE RESEARCH DIRECTIONS

UNDERSTANDING SOCIAL, CULTURAL, AND INDIVIDUAL PREFERENCES FROM DATA

Social and cultural backgrounds can affect one’s judgment of aesthetics or influence one’s emotions in a particular

scenario. An important future research direction would be to incorporate cultural, social, and personal differences into the learning methodologies. An important starting point can be to determine how many distinct “preference groups” (cultural or social) there are in a population. This could be followed by discovering characteristic rating distributions of scores that differ across different preference groups. Semantics can also play a role in aesthetics or emotional judgments, especially as perception of semantics may vary across cultures. Through these and related questions, an attempt can be made to understand the relationships between individuals, preference groups, and masses. While consensus measures and averaged-out ratings provide a generic learning setting, personalized models are of high relevance because of the significant amount of subjectivity in the problems and therefore may be valuable for practical applications.

One can explore personalization at two different levels. First, one can consider preference groups or cliques within a given context, i.e., groups of people who share similar tastes within a social or cultural setting, followed by understanding tastes of individuals. The main problem with the latter is that a significant amount of personalized data is needed to learn a reasonable model for each individual, which is typically not available. If there indeed is a finite set of cliques in the population, then clique-specific models can be learned. One can follow this path by discovering the cliques in the population and learning clique-specific models. If one treats the clique membership of individuals as a soft assignment, whereby each person belongs to different cliques with a certain probability, one obtains a simple model of personalization as well. Personalization paradigms can be explored, drawing inspiration from the collaborative and content-based filtering literature [6], [38].

UNDERSTANDING PSYCHOLOGY-INDUCED DIFFERENCE IN JUDGMENTS

Emotional and aesthetic impact of art and visual imagery is also linked to the emotional state of the viewer, who, according to the emotional congruence theory, perceives his or her environment in a manner congruent with his/her current emotional state. The latter is based on emotional-congruent or mood-congruent processing where a person’s mood can sensitize the person to take in mainly information that agrees with his/her mood [15]. Studies have shown that art preferences and art judgment can vary significantly across expert and nonexpert subjects. For example, there is a higher correlation between originality and quality of art for experts than for nonexperts [43] and experts accord more value to originality in determining aesthetic value. Furthermore, artists and experienced art viewers tend to prefer artworks that are challenging and emotionally provocative [98], which is in contrast to the majority of people who prefer art that makes them happy and feel relaxed [100]. All of these findings emphasize the necessity to consider individual differences in experimental research on aesthetics. The results reported in [5], [16], and [33] demonstrate that these differences are significant and can be explained on the basis of common mechanisms as suggested by Berlyne in [9].

UNDERSTANDING AND MODELING CONTEXT

Context plays an important role in semantic image understanding [64]. Context within the purview of images has been explored as spatial context (leveraging spatial arrangement of objects in images), temporal context (leveraging the time and date information when pictures were taken), geographical context (leveraging information about geographical location of pictures) [48], [49], and social context [37], [79], [105] (leveraging information about the social circle of a person or social relationship reflected in pictures). For example, people may well associate special emotions with pictures taken on special occasions or about special people in their lives. Similarly, pictures taken during one’s trip to a national park may be aesthetically more pleasing than pictures taken in a local park, purely because of their content and opportunities for high-quality shots. Determining the extent to which such factors affect the aesthetic or emotional value of pictures will be a potent future research direction. At the same time, the nature of the data being used for a specific problem can largely influence aesthetics or emotional models. In truth, none of the models proposed will be fundamental or absolute in what they learn about aesthetics or emotions, but will be tempered to the given data acquisition step. For example, what is considered “interesting” (Flickr) may not be treated as being “aesthetically pleasing” (Photo.net) by the population, and vice versa. Examples of key contextual aspects of data are a) the exact question posed to the users about the images, e.g., “aesthetics” [115], “overall quality” [111], or “like it” [110]; b) the type of people who visit and vote on the images, e.g., general enthusiasts [115], [111] or photographers [115]; and c) the type of images rated, e.g., travel [116] or topical [111]. One long-term goal would be to look for solutions that apply to as general a context as possible.

ATTEMPTING BRUTE FORCE DATA-DRIVEN APPROACHES

The World Wide Web is growing at a phenomenal rate and so is the amount of image data in Web-based photo sharing repositories such as Flickr. This is evident by the fact that on an average about 5,000 images are uploaded to Flickr every minute. The availability and potential usability of Web users as information providers has been leveraged by some researchers to design games that prompt users to provide tags and other metadata for images [1], [2]. While they provide a source for enjoyment, the games also have a deeper goal to collect high-quality metadata that is expected to greatly complement visual search. In the wake of this, an interesting diversion from sophisticatedly crafted algorithms is to explore brute force methods for image understanding tasks [36], [41], [42], [89], [95]. Such approaches employ simplistic search methods into massive repositories of image data to achieve recognition. The basic philosophy behind these brute force search techniques is that the content and attributes of a query image can be collectively inferred from visually similar pictures, and the inference is expected to improve as the size of the search-space grows. While it remains an open problem as to how large-scale, data-driven methods would perform in the task of aesthetics or emotion recognition, the success of the brute force philosophy in several recognition

tasks provides hope that this could be an interesting future research area for the problems discussed in this article.

DEVELOPING REAL-WORLD USABLE RESEARCH PROTOTYPES

Perhaps one of the most important steps in the life cycle of a research idea is its incorporation into a usable and testable system open to the scrutiny of common people. This is important for two reasons: 1) it provides a realistic test bed for evaluating the research machinery, and 2) user reaction and feedback can be very useful in helping the design of future prototypes. In light of this, a key future direction could be to take some of the proposed ideas in the current research domain to the next level in their life cycle. We briefly describe ACQUINE [109], an attempt in this direction. ACQUINE is a machine-learning-based online system that showcases computer-based prediction of aesthetic quality for color natural photographic pictures (Figure 1). Labeled images from Photo.net have been obtained to achieve supervised learning of aesthetic quality rating models. A number of visual features that are assumed to be correlated with aesthetic quality are extracted from images and an SVM-based classifier is used to obtain the aesthetic rating of a given picture. Users can upload their own images, use links to images that exist on the Web, or simply browse photographs uploaded by others. They are also able to look at the ratings that were machine-given, and optionally add their own rating. This is a valuable source of feedback and labeled data for future iterations of the system. As of May 2011, nearly 250,000 images from nearly 32,000 different users have been uploaded to ACQUINE for automatic rating. Over 65,000 user ratings of photos have also been provided.

In this tutorial, we have looked at key aspects of aesthetics, emotions, and associated computational problems with respect to natural images and artwork. We discussed these problems in relation to philosophy, photography, paintings, visual arts, and psychology. Computational frameworks and representative approaches proposed to address problems in this domain were outlined followed by a discussion of available data sets for research use. An analysis of the nature of data and ratings among the available resources was also presented. In conclusion, we laid out a few intriguing directions for future research in this area. We hope that this tutorial will significantly increase the visibility of this research area and foster dialogue and collaboration among artists, photographers, and researchers in signal processing, computer vision, pattern recognition, and psychology.

ACKNOWLEDGMENTS

The authors acknowledge the constructive comments of the anonymous reviewers. J.Z. Wang and J. Li would like to thank the Van Gogh and Kröller-Müller Museums for providing the photographs of paintings for their study. National Science Foundation Grants IIS-0347148, CCF-0936948, and EIA-0202007 provided partial funding for their research. Part of the work of J.Z. Wang was done while working at the National Science Foundation. Any opinions, findings, and conclusions or recommendations

expressed in this material are those of the authors and do not necessarily reflect the views of the Foundation.

AUTHORS

Dhiraj Joshi (dhiraj.joshi@kodak.com) is a principal scientist at Eastman Kodak Research Labs. His primary focus is association of image content, tags, and location metadata with image semantics. He also works on intelligent systems that use semantics across multiple modalities of media for an enriched user experience. He has previously worked on large-scale image retrieval systems. He was a research intern at IBM T.J. Watson Research Labs and the IDIAP Research Institute (Switzerland). In 2006, he was selected by IBM Research as an emerging leader in multimedia research to present at the Watson Emerging Leaders in Multimedia Workshop. He is a Member of the IEEE and currently serves as the chair of the Rochester Chapter of the IEEE Signal Processing Society.

Ritendra Datta (ritendra@google.com) is an engineer at Google. He completed his Ph.D. degree from Penn State University in 2009. His research interests include statistical modeling and machine learning, image content analysis, content-based image search, automatic image tagging, visual aesthetic quality inference, and social networks. He was a research intern at IBM T.J. Watson Research Center, Xerox PARC, and Google. He was a recipient of the Glenn Singley Memorial Graduate Fellowship in Engineering at Penn State in 2004 and was invited as an Emerging Leader in Multimedia by IBM Research in 2007.

Elena Fedorovskaya (elena.fedorovskaya@kodak.com) has been a research scientist at Kodak Research Laboratories since 1997. She received a Ph.D. degree in psychophysiology and an M.Sc. degree in applied mathematics, both from Lomonosov Moscow State University (Russia). She worked as a research scientist from 1986 to 1997 at the Department of Psychophysiology, Moscow Lomonosov State University, conducting research in the area of psychophysiology of vision and cognition and electrophysiology of stress. She also visited the Institute for Perception Research at the Technical University of Eindhoven, The Netherlands to study image perception and image quality. At Kodak, she works on understanding and modeling perceptual, cognitive, and emotional aspects of human experience in relation to images and imaging systems.

Quang-Tuan Luong (QTL@terrageria.com) is a full-time freelance nature and travel photographer from San Jose, California. He received his Ph.D degree from the University of Paris (Orsay). When he came to the United States to conduct research in the fields of artificial intelligence and image processing, he fell in love with the national parks. After he became the first to photograph all of them in large format, Ken Burns featured him in "The National Parks: America's Best Idea" (2009). His photographs have been the subject of two coffee-table books and have been published in three dozen countries. They have been profiled in several photography magazines and *National Geographic Explorer*, as well as seen in galleries and museum exhibitions on both U.S. coasts.

James Z. Wang (jwang@ist.psu.edu) received a bachelor's degree (summa cum laude) from the University of Minnesota, an M.Sc. degree in mathematics, an M.Sc. degree in computer science, and a Ph.D. degree in medical information sciences from Stanford University. He is a program manager in the Office of International Science and Engineering, Office of the Director, the U.S. National Science Foundation. He has been on the faculty at the College of Information Sciences and Technology of Penn State since 2000, where he is a professor. His research interests are image database retrieval, computational aesthetics, image tagging, climate informatics, and biomedical informatics. He received an NSF Career Award and the endowed PNC Technologies Career Development Professorship.

Jia Li (jiali@stat.psu.edu) is an associate professor of statistics and (by courtesy appointment) in computer science and engineering at The Pennsylvania State University, University Park. She received the M.Sc. degree in electrical engineering, the M.Sc. degree in statistics, and the Ph.D. degree in electrical engineering from Stanford University. She was a visiting scientist at Google Labs in Pittsburgh (2007–2008), a research associate in the Computer Science Department at Stanford University (1999), and a researcher at the Xerox Palo Alto Research Center (1999–2000). Her research interests include statistical modeling and learning, data mining, computational biology, image processing, and image annotation and retrieval.

Jiebo Luo (jiebo.luo@kodak.com) is a senior principal scientist with the Kodak Research Laboratories, Rochester, New York. His research interests include image processing, machine learning, computer vision, computational photography, biomedical imaging and informatics, multimedia data mining, and ubiquitous computing. He has authored more than 160 technical papers and holds over 60 issued U.S. patents. He is the editor-in-chief of the *Journal of Multimedia*. He also serves on the editorial boards of *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *IEEE Transactions on Multimedia*, and *IEEE Transactions on Circuits and Systems for Video Technology, Pattern Recognition, and Machine Vision and Applications*. He has been involved in organizing numerous leading technical conferences sponsored by the IEEE, ACM, and SPIE. He is a Fellow of the IEEE, SPIE, and IAPR.

REFERENCES

[1] L. von Ahn, "Games with a purpose," *IEEE Comput. Mag.*, vol. 39, no. 6, pp. 98–98, June 2006.

[2] L. von Ahn and L. Dabbish, "Designing games with a purpose," *Commun. ACM*, vol. 51, no. 8, pp. 58–67, Aug. 2008.

[3] I. Arapakis, I. Konstas, and J. M. Jose, "Using facial expressions and peripheral physiological signals as implicit indicators of topical relevance," in *Proc. ACM Multimedia*, 2009, pp. 461–470.

[4] R. Arnheim, *Art and Visual Perception: A Psychology of the Creative Eye*. Los Angeles: Univ. California Press, 1954.

[5] O. Axelsson, "Towards a psychology of photography: Dimensions underlying aesthetic appeal of photographs," *Percept. Mot. Skills*, vol. 105, no. 2, pp. 411–434, 2007.

[6] M. Balabanovic and Y. Shoham, "Fab: Content-based, collaborative recommendation," *Commun. ACM*, vol. 40, no. 3, pp. 66–72, 1997.

[7] B. Barry, "The mindful camera: Common sense for documentary videography," in *Proc. ACM Multimedia*, 2003, pp. 648–649.

[8] B. Barry and G. Davenport, "Documenting life: Videography and common sense," in *Proc. IEEE Int. Conf. Multimedia (ICME)*, 2003, pp. 197–200.

[9] D. E. Berlyne, *Aesthetics and Psychobiology*. New York: Appleton-Century-Crofts, 1971.

[10] N. Bianchi-Berthouze, "K-dime: An affective image filtering system," *IEEE Multimedia*, vol. 10, no. 3, pp.103–106, 2003.

[11] S. Bhattacharya, R. Sukthankar, and M. Shah, "A framework for photo-quality assessment and enhancement based on visual aesthetics," in *Proc. ACM Multimedia*, 2010, pp. 271–280.

[12] M. Bressan, C. Cifarelli, and F. Perronnin, "An analysis of the relationship between painters based on their work," in *Proc. IEEE ICIP*, 2008, pp. 113–116.

[13] I. E. Bereznoy, E. O. Postma, and H. J. van den Herik, "Computerized visual analysis of paintings," in *Proc. 16th Int. Conf. Association History and Computing*, 2005, pp. 28–32.

[14] I. E. Bereznoy, E. O. Postma, and H. J. van den Herik, "Computer analysis of Van Gogh's complementary colors," *Pattern Recognit. Lett.*, vol. 28, no. 6, pp. 703–709, 2007.

[15] G. H. Bower, "Mood and memory," *Amer. Psychol.*, vol. 36, no. 2, pp. 129–148, Feb. 1981.

[16] C. Cerosaletti and A. Loui, "Measuring the perceived aesthetic quality of photographic images," in *Proc. 1st Int. Workshop Quality Multimedia Experience*, 2009, pp. 47–52.

[17] B. Cheng, B. Ni, S. Yan, and Q. Tian, "Learning to photograph," in *Proc. ACM Multimedia*, 2010, pp. 291–300.

[18] S. Daly, "The visible differences predictor: An algorithm for the assessment of image fidelity," in *Digital Image Hum. Vis.* Cambridge, MA: MIT Press, 1993, pp. 179–206.

[19] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in *Proc. ECCV*, 2006, pp. 288–301.

[20] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Comput. Surv.*, vol. 40, no. 2, pp. 51–60, 2008.

[21] R. Datta, J. Li, and J. Z. Wang, "Learning the consensus on visual quality for next generation image management," in *Proc. ACM Multimedia*, 2007, pp. 533–536.

[22] R. Datta, J. Li, and J. Z. Wang, "Algorithmic inferring of aesthetics and emotion in natural images: An exposition," in *Proc. ICIP*, 2008, pp. 105–108.

[23] B. C. Davis and S. Lazebnik, "Analysis of human attractiveness using manifold kernel regression," in *Proc. ICIP*, 2008, pp. 109–112.

[24] J. O'Doherty, J. Winston, H. Critchley, D. Perrett, D. M. Burt, and R. J. Dolan, "Beauty in a smile: The role of medial orbitofrontal cortex in facial attractiveness," *Neuropsychologia*, vol. 41, no. 2, pp. 147–155, 2003.

[25] D. Dutton, *The Art Instinct: Beauty, Pleasure, and Human Evolution*. New York, NY: Bloomsbury Press, 2009.

[26] J. Elkins, *The Domain of Images*. Ithaca, NY: Cornell Univ. Press, 1999.

[27] J. Elkins, "Aesthetics and the two cultures: Why art and science should be allowed to go their separate ways," in *Rediscovering Aesthetics: Transdisciplinary Voices from Art History, Philosophy, and Art Practice (Cultural Memory in the Present)*, F. Halsall, J. Jansen, and T. O'Connor, Eds. New York: Columbia Univ. Press, 2009, pp. 34–50.

[28] Y. Eysenck, G. Dror, and E. Ruppin, "Facial attractiveness: Beauty and the machine," *Neural Comput.*, vol. 18, no. 1, pp. 119–142, 2006.

[29] C. M. Falco, "Computer vision and art," *IEEE Multimedia*, vol. 14, no. 2, pp. 8–11, 2007.

[30] Y. Fang, D. Geman, and N. Boujemaa, "An interactive system for mental face retrieval," in *Proc. ACM SIGMM Int. Workshop Multimedia Information Retrieval*, 2005, pp. 193–200.

[31] G. T. Fechner, "Zur experimentalen Ästhetik (On experimental aesthetics)," *Abhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften*, vol. 9, pp. 555–635, 1871.

[32] G. T. Fechner, "Vorschule der ästhetik," *Breitkopf und Härtel, Leipzig*, Breitkopf und Härtel, Leipzig, vols. 1–2, 1876.

[33] E. A. Fedorovskaya, C. Neustaedter, and W. Hao, "Image harmony for consumer images," in *Proc. ICIP*, 2008, pp. 121–124.

[34] M. Freeman, *The Photographer's Eye: Composition and Design for Better Digital Photos*. Waltham, MA: Focal Press, Elsevier Inc., 2007.

[35] S. Freud, *The Interpretation of Dreams*. New York: The Macmillan Company, 1913.

[36] A. Gallagher, D. Joshi, J. Yu, and J. Luo, "Geo-location inference from image content and user tags," in *Proc. IEEE Int. Workshop Internet Vision (CVPR)*, 2009, pp. 55–62.

[37] A. Gallagher and T. Chen, "Using context to recognize people in consumer images," *IPSI Trans. Comput. Vis. Applicat.*, vol. 1, pp. 115–126, March 2009.

[38] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Commun. ACM*, vol. 35, no. 12, pp. 61–70, 1992.

[39] N. Goodman, *Languages of Art: An Approach to a Theory of Symbols*, 2nd ed. Indianapolis, IN: Hackett Publishing Co., 1976.

[40] C. Greenberg, *Art and Culture Critical Essays*. Boston, MA: Beacon Press, 1971.

[41] J. Hays and A. Efros, "Scene completion using millions of photographs," *ACM Trans. Graphics*, vol. 26, no. 2, 2007.

[42] J. Hays and A. Efros, "IM2GPS: Estimating geographic information from a single image," in *Proc. CVPR*, 2008, pp. 1–8.

[43] P. Hekkert, "Beauty in the eye of expert and non-expert beholders: A study in the appraisal of art," *Amer. J. Psychol.*, vol. 109, no. 3, pp. 389–407, 1997.

[44] J. Howe, "The rise of crowdsourcing," *Wired Mag.*, vol. 14, no. 6, June 2006.

[45] C. R. Johnson, Jr., E. Hendriks, I. J. Bereznoy, E. Brevdo, S. M. Hughes, I. Daubechies, J. Li, E. Postma, and J. Z. Wang, "Image processing for artist

- identification: computerized analysis of Vincent van Gogh's painting brushstrokes," *IEEE Signal Processing Mag. (Special Issue on Visual Cultural Heritage)*, vol. 25, no. 4, pp. 37–48, 2008.
- [46] H. Kawabata and S. Zeki, "Neural correlates of beauty," *J. Neurophysiol.*, vol. 91, no. 4, pp. 1699–1705, April 2004.
- [47] Y. Ke, X. Tang, and F. Jing, "The design of high-level features for photo quality assessment," in *Proc. CVPR*, 2006, pp. 419–426.
- [48] L. Kennedy and M. Naaman, "How flicker helps us make sense of the world: Context and content in community-contributed media collections," in *Proc. ACM Multimedia*, 2007, pp. 631–640.
- [49] L. Kennedy and M. Naaman, "Generating diverse and representative image search results for landmarks," in *Proc. 17th Int. Conf. World Wide Web*, 2008, pp. 297–306.
- [50] U. Kirk, M. Skov, O. Hulme, M. S. Christensen, and S. Zeki, "Modulation of aesthetic value by semantic context: An fMRI study," *Neuroimage*, vol. 44, no. 3, pp. 1125–1132, Feb 2009.
- [51] K. Koffka, *Gestalt Psychology*. Orlando, FL: Harcourt Brace Jovanovic, 1935.
- [52] S. Kroner and A. Lattner, "Authentication of free hand drawings by pattern recognition methods," in *Proc. IEEE ICPR*, 1998, pp. 462–464.
- [53] A. Kushki, P. Androustos, K. Plataniotis, and A. Venetsanopoulos, "Retrieval of images from artistic repositories using a decision fusion framework," *IEEE Trans. Image Process.*, vol. 13, no. 3, pp. 277–292, 2004.
- [54] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm, "Looking at pictures: Affective, facial, visceral, and behavioral reactions," *Psychophysiology*, vol. 30, no. 3, pp. 261–273, May 1993.
- [55] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): Technical, manual, and affective ratings," NIMH Center for the Study of Emotion and Attention, Gainesville, FL, 1997.
- [56] R. Lattó, "The brain of the beholder," in *The Artful Eye*. New York, NY: Oxford Univ. Press, 1995, pp. 66–94.
- [57] C. C. Li and T. Chen, "Aesthetic visual quality assessment of paintings," *IEEE J. Select. Topics Signal Process.*, vol. 3, no. 2, pp. 236–252, 2009.
- [58] J. Li and J. Z. Wang, "Studying digital imagery of ancient paintings by mixtures of stochastic models," *IEEE Trans. Image Process.*, vol. 13, no. 3, pp. 340–353, 2004.
- [59] Y. Liu, K. L. Schmidt, J. F. Cohn, and S. Mitra, "Facial asymmetry quantification for expression invariant human identification," in *Proc. CVPR*, 2003, pp. 198–204.
- [60] A. Louis, M. D. Wood, A. Scalise, and J. Birkelund, "Multidimensional image value assessment and rating for automated albuming and retrieval," in *Proc. ICIP*, 2008, pp. 97–100.
- [61] P. J. Lu and P. J. Steinhardt, "Decagonal and quasi-crystalline tilings in medieval Islamic architecture," *Science*, vol. 315, no. 5815, pp. 1106–1110, 23 Feb 2007.
- [62] P. J. Lu, "Early precision compound machine from ancient China," *Science*, vol. 304, no. 5677, p. 1638, 11 June 2004.
- [63] J. Luo, A. Savakis, S. Etz, and A. Singhal, "On the application of Bayes networks to semantic understanding of consumer photographs," in *Proc. ICIP*, 2000, pp. 512–515.
- [64] J. Luo, M. Boutell, and C. Brown, "Exploiting context for semantic scene content understanding," *IEEE Signal Processing Mag. (Special Issue on Semantic Retrieval of Multimedia)*, vol. 23, no. 2, pp. 101–114, 2006.
- [65] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," in *Proc. ACM Multimedia*, 2010, pp. 83–92.
- [66] W. J. T. Mitchell, *Iconology: Image Text and Ideology*. Chicago, IL: Univ. Chicago Press, 1986.
- [67] W. J. T. Mitchell, *Picture Theory*. Chicago, IL: Univ. Chicago Press, 1994.
- [68] C. F. Nodine, P. J. Locher, and E. A. Krupinski, "The role of formal art training on perception and aesthetic judgment of art compositions," *Leonardo*, vol. 26, no. 3, pp. 219–227, 1993.
- [69] S. E. Palmer, "Aesthetic science: Human preferences for spatial composition," Keynote address in *Proc. IS&T/SPIE Electronic Imaging Conf.*, 2009.
- [70] D. I. Perrett, K. A. May, and S. Yoshikawa, "Facial shape and judgments of female attractiveness," *Nature*, vol. 368, pp. 239–242, 17 March 1994.
- [71] G. Peters, "Aesthetic primitives of images for visualization," in *Proc. IEEE Int. Conf. Information Visualization*, 2007, pp. 316–325.
- [72] V. S. Ramachandran and W. Hirstein, "Science of art: A neurological theory of aesthetic experience," *J. Consciousness Stud.*, vol. 6, no. 6/7, pp. 15–51, June/July 1999.
- [73] S. Ramanathan, H. Katti, R. Huang, T.-S. Chua, and M. Kankanalli, "Automated localization of affective objects and actions in images via caption text-cum-eye gaze analysis," in *Proc. ACM Multimedia*, 2009, pp. 729–732.
- [74] R. N. Reber, N. Schwartz, and P. Winkielman, "Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?" *Pers. Social Psychol. Rev.*, vol. 8, no. 4, pp. 364–382, 2004.
- [75] D. Rockmore, S. Lyu, and H. Farid, "A digital technique for authentication in the visual arts," *Int. Found. Art Res.*, vol. 8, no. 2, pp. 12–23, 2006.
- [76] A. Savakis, S. Etz, and A. Loui, "Evaluation of image appeal in consumer photography," in *Proc. SPIE Human Vision and Electronic Imaging*, 2000, pp. 111–120.
- [77] J. E. Scheib, S. W. Gangestad, and R. Thornhill, "Facial attractiveness, symmetry, and cues of good genes," *Proc. Royal Soc. London, Biol Sci.*, vol. 266, no. 1431, pp. 1913–1917, 22 Sept 1999.
- [78] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Processing*, vol. 14, no. 11, pp. 1918–1927, 2005.
- [79] B. Shevade, H. Sundaram, and L. Xie, "Modeling personal and social network context for event annotation in images," in *Proc. Joint Conf. Digital Libraries*, 2007, pp. 127–134.
- [80] P. Singh and B. Barry, "Teaching machines about everyday life," *BT Technol. J.*, vol. 22, no. 4, pp. 227–240, 2004.
- [81] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of early years," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 12, pp. 1349–1380, 2000.
- [82] R. L. Solso, *The Psychology of Art and the Evolution of the Conscious Brain*. Cambridge, MA: MIT Press, 2003.
- [83] I. Stamos and P. Allen, "3-D model construction using range and image data," in *Proc. CVPR*, 2000, pp. 531–536.
- [84] D. Stork, "Computer vision and computer graphics analysis of paintings and drawings: An introduction to the literature," in *Proc. Int. Conf. Computer Analysis of Images and Patterns (LNCS 5702)*. Berlin: Springer-Verlag, 2009, pp. 9–24.
- [85] X. Sun, H. Yao, R. Ji, and S. Liu, "Photo assessment based on computational visual attention model," in *Proc. ACM Multimedia*, 2009, pp. 541–544.
- [86] J. P. Swaddle and I. C. Cuthill, "Asymmetry and human facial attractiveness: Symmetry may not always be beautiful," in *Proc. Royal Soc. London, Biol Sci.*, vol. 261, no. 1360, pp. 111–116, 2 July 1995.
- [87] R. Taylor, A. P. Micolich, and D. Jones, "Fractal analysis of pollock's drip paintings," *Nature*, vol. 399, no. 6735, p. 422, 1999.
- [88] R. Taylor, "Pollock, Mondrian and the nature: Recent scientific investigations," *Chaos Complexity Lett.*, vol. 1, no. 3, pp. 265–277, 2004.
- [89] A. Torralba, R. Fergus, and W. T. Freeman, "80 million tiny images: A large dataset for nonparametric object and scene recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 30, no. 11, pp. 1958–1970, 2008.
- [90] R. S. Ulrich and L. Gilpin, "Healing arts: Nutrition for the soul," in *Putting Patients First: Designing and Practicing Patient-Centered Care*. San Francisco: Jossey-Bass (Wiley), 2003, pp. 117–146.
- [91] R. Valenti, N. Sebe, and T. Gevers, "Facial expression recognition: A fully integrated approach," in *Proc. Int. Workshop Visual and Multimedia Digital Libraries*, 2007, pp. 125–130.
- [92] R. Valenti, A. Jaimes, and N. Sebe, "Sonify your face: Facial expressions for sound generation," in *Proc. ACM Multimedia*, 2010, pp. 1363–1372.
- [93] C. W. Valentine, *The Experimental Psychology of Beauty*. London: Methuen and Co. Ltd Publishers, 1962.
- [94] W. Wang and Q. He, "A survey on emotional semantic image retrieval," in *Proc. ICIP*, 2008, pp. 117–120.
- [95] X. J. Wang, L. Zhang, F. Jing, and W. Y. Ma, "Annosearch: Image auto-annotation by search," in *Proc. CVPR*, 2006, pp. 1483–1490.
- [96] A. B. Watson, "Toward a perceptual video quality metric," *Proc. SPIE*, vol. 3299, pp. 139–147, January 1998.
- [97] A. R. Willis and D. B. Cooper, "Computational reconstruction of ancient artifacts—From ruins to relics," *IEEE Signal Processing Mag.*, vol. 25, no. 4, pp. 65–83, 2008.
- [98] A. S. Winston and G. C. Cupchik, "The evaluation of high art and popular art by naive and experienced viewers," *Vis. Arts Res.*, vol. 18, no. 1, pp. 1–14, 1992.
- [99] R. Wollheim, *Painting as an Art*. Princeton, NJ: Princeton Univ. Press, 1987.
- [100] J. Wypijewski, Ed., *Painting by the Numbers: Komar and Melamid's Scientific Guide to Art*. New York: Farrar, Straus and Giroux, 1997.
- [101] Y. Yang, M. Song, N. Li, J. Bu, and C. Chen, "Visual attention analysis by pseudo gravitational field," in *Proc. ACM Multimedia*, 2010, pp. 553–556.
- [102] V. Yanulevskaya, J. C. van Gemert, K. Roth, A. K. Herbold, N. Sebe, and J. M. Geusebroek, "Emotional valence categorization using holistic image features," in *Proc. ICIP*, 2008, pp. 101–104.
- [103] D. W. Zaidel and J. A. Cohen, "The face, beauty, and symmetry: Perceiving asymmetry in beautiful faces," *Int. J. Neurosci.*, vol. 115, no. 8, pp. 1165–1173, August 2005.
- [104] S. Zeki, *Inner Vision: An Exploration of Art and the Brain*. New York: Oxford Univ. Press, 1999.
- [105] A. Zunjarwad, H. Sundaram, and L. Xie, "Contextual wisdom: Social relations and correlations for multimedia event annotation," in *Proc. ACM Multimedia*, 2007, pp. 615–624.
- [106] "Special issue on image processing for cultural heritage," *IEEE Trans. Image Processing*, vol. 13, no. 3, 2004.
- [107] "Special issue on semantic retrieval of multimedia," *IEEE Signal Processing Mag.*, vol. 23, no. 2, 2006.
- [108] "Special issue on visual cultural heritage," *IEEE Signal Processing Mag.*, vol. 25, no. 4, 2008.
- [109] ACQUINE [Online]. Available: <http://acquine.alipr.com>
- [110] ALIPR [Online]. Available: <http://alipr.com>
- [111] DPChallenge [Online]. Available: <http://www.dpchallenge.com>
- [112] Encyclopedia Britannica [Online]. Available: <http://www.britannica.com>
- [113] Flickr [Online]. Available: <http://www.flickr.com>
- [114] Nadia Camera [Online]. Available: <http://www.wired.com/gadgetlab/2010/07/nadia-camera-offers-opinion-of-your-terrible-photos/>
- [115] Photo.net [Online]. Available: <http://photo.net>
- [116] Terragalleria [Online]. Available: <http://www.terragalleria.com>
- [117] USA Today [Online]. Available: http://www.usatoday.com/tech/news/technovations/2006-12-18-computer-feelings_x.htm