

On Aesthetics and Emotions in Images:

A Computational Perspective

Dhiraj Joshi, Ritendra Datta, Elena Fedorovskaya, Xin Lu, Quang-Tuan Luong,

James Z. Wang, Jia Li, Jiebo Luo

Abstract - In this chapter, we discuss the problem of computational inference of aesthetics and emotions from images. We draw inspiration from diverse disciplines such as philosophy, photography, art, and psychology to define and understand the key concepts of aesthetics and emotions. We introduce the primary computational problems that the research community has been striving to solve and the computational framework required for solving them. We also describe datasets available for performing assessment and outline several real-world applications where research in this domain can be employed. This chapter discusses the contributions of a significant number of research articles that have attempted to solve problems in aesthetics and emotion inference in the last several years. We conclude the chapter with directions for future research.

I. INTRODUCTION

The image processing community together with vision and computer scientists have, for a long time, attempted to solve image quality assessment [67][34][12][81] and image semantics inference [14]. 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 [13][15][77][78].

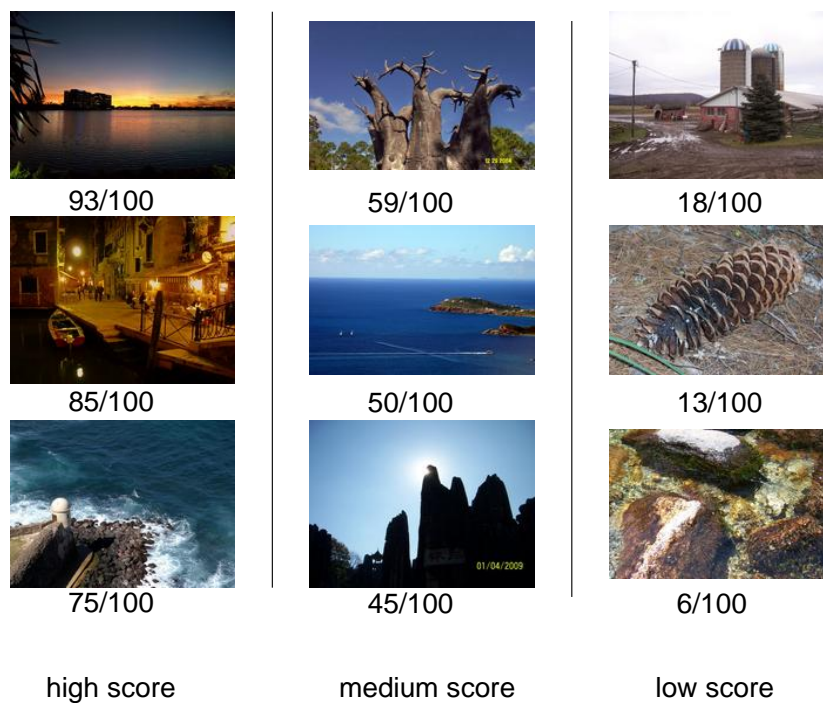


Figure 1: Pictures with high, medium, and low aesthetics scores from ACQUINE, an online automatic photo aesthetics engine.

Fig. 1 shows an example of state-of-the-art automatic aesthetics assessment. 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 [41][58] and aesthetics in visual art, including photography, are also natural and essential.

Despite the challenges, various research attempts have been made and are increasingly being made to address basic understanding and solve various sub-problems under the umbrella of aesthetics, mood, and emotion inference in pictures. 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

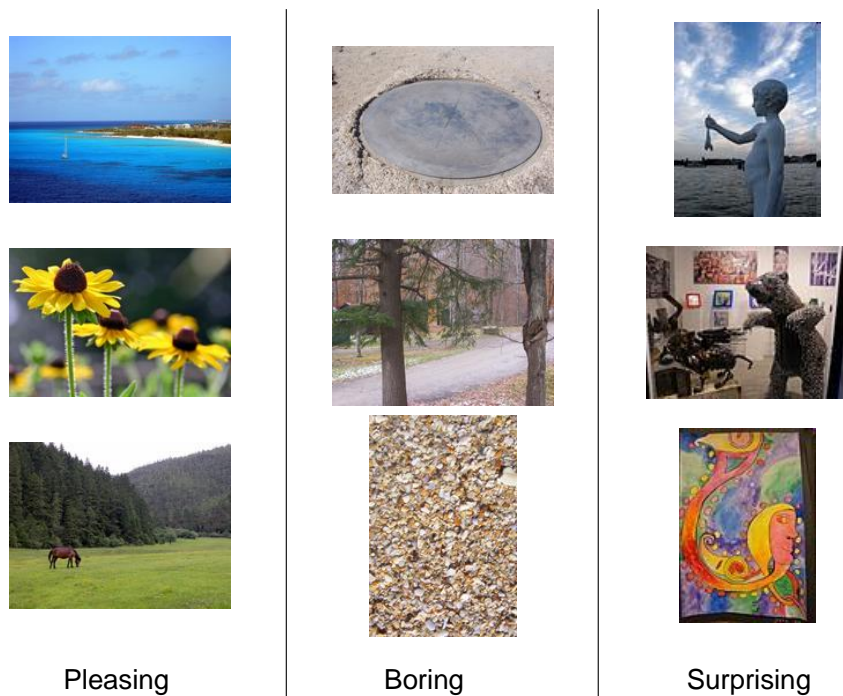


Figure 2: Pictures and emotions rated by users from ALIPR.com, a research site for machine-assisted image tagging.

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. Picture editors and photographers can make use of automated aesthetics feedback when selecting photos for photo-clubs, competitions, portfolio reviews, or workshops. Similarly, from a publication perspective, a museum curator may be interested in assessing if an artwork is enjoyable by a majority of the people. 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 such expert decision making and perhaps with time and feedback learn to adapt to expert opinion better (Fig. 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 [71]).

In this chapter, we have attempted to introduce components that are essential for the broader research community to get involved and excited about this field of study. In Section II, we discuss aesthetics with respect to philosophy, photography, art, and psychology. Section III introduces a wide spectrum of research problems that have been attempted in computational aesthetics and emotions. The computational framework in the form of feature extraction, representation, and modeling is the topic of Section IV. Datasets and other resources available for aesthetics and emotions research are reviewed in Section V while Section VI takes a futuristic stance and discusses potential research directions and applications.

II. 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 have resulted in formation of two views on beauty and aesthetics: 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.

A. 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. For several years, Photo.net has been a place for photographers to rate the photos of peers [96]. Here a photo is rated along two dimensions, aesthetics and originality, each with a score between one and seven. 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.”

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. Photographers used techniques such as soft focus, special filters, lens coatings, special darkroom processing, and printing to achieve desired artistic effects in their pictures. 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. Post-modernism rejected ideas of objective truth in art. Sharp classifications into high-art and low-art became defunct.

In spite of these differing 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. One such accepted rule being that 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 again common and not-so-common theories or rules. The rule of thirds is the most widely known which 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. 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. 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.”

B. 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. 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.



Figure 3: Paintings by Van Gogh (top-left) Avenue of Poplars in Autumn, (top right) Still Life: Vase with Gladioli, (bottom-left) Willows at Sunset, (bottom-right) Automatically extracted brushstrokes for Willows at Sunset. Notice the widely different nature and use of colors in the paintings (courtesy – Top images: Van Gogh Museum Amsterdam (Vincent van Gogh Foundation). Bottom images: Kröller-Müller Museum and James Z. Wang Research Group at Penn State.).

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. One of the key movements of Western art, Impressionism, started in late 19th century with Claude Monet's masterpiece "Impression, Sunrise, 1872." Impressionist artists focused on ordinary subject matter, painted outdoors, used visible brush-strokes, 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. 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 in order to express his artistic ideas (Fig. 3). Van Gogh also developed a bold style of brush strokes, an understanding of which can perhaps offer newer perspectives into understanding his work and that of his contemporaries (Fig. 3 shows an example of automatic brushstroke extraction research presented in [32]).

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 and possibilities of creation became virtually unbounded. Although there has recently been some work on inferring aesthetics in paintings [44][75][76], such work is usually limited to a small-scale specific experimental setup. One such work [76] scientifically examines the works of Mondrian and Pollock, two stalwarts of modern art with drastically distinct styles (the former attempting to achieve spiritual harmony in art while the latter known for mixing sand, broken glass, and paint and his unconventional paint drip technique).

C. 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 Aesthetick* in 1876 [23][24]. Fechner suggested three methods for use in experimental aesthetics, (i) including the method of choice where subjects are asked to compare objects with respect to their pleasingness; (ii) the method of production, where subjects are required to produce an object that conforms to their tastes by drawing or other actions; and (iii) 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.

Developments in other areas of psychology of the early decades of the twentieth century contributed to the psychology of aesthetics. Gestalt psychology produced influential ideas such as the concept of goodness of patterns and configurations emphasizing regularity, symmetry, simplicity, and closure [38]. 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) [3], Berlyne formulated several theoretically and experimentally substantiated ideas that helped shape modern experimental research in aesthetics into the science of aesthetics [57].

Berlyne's ideas and research directions together with the advances in understanding of neural mechanisms of perception, cognition, and emotion obtained in psychology [70], psychophysiology, and neuroscience and facilitated by the modern imaging techniques led to the emergence of neuroaesthetics in the 1990s [33][37][60][89]. Recent studies associated with the Processing Fluency Theory by Reber et al. in [62] 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.

III. 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 people. 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.

A. Core Problems

1) Aesthetics Prediction

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 [13] where the quality of score prediction is assessed in the form of rate or distribution of error.

It has been observed both in [13] and [34] that score prediction is a highly challenging problem, mainly due to noise in user ratings. To make the problem more solvable, the regression problem is changed to one of classification, by thresholding the average scores to create high- vs. low-quality image classes [13], or professional vs. snapshot image classes [34]. 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 [15] has proven to be more successful than the general classification problem. The classification problem solutions can be evaluated by standard accuracy measures [13][34]. 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: An aesthetics 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 multi-class classification problem well known in machine learning). Opportunities lie in learning and predicting “distributions of aesthetics values” instead of singular aesthetics classes or scores. 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. An effort in this direction has been made in [83]

2) 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 classification problem [86]. Consider that there are K 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., $k(I) = \mathit{argmax}_i \Pi_i(I)$. The problem is only meaningful when there is clear dominance of $k(I)$ over others.

Emotion Distribution: 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 a 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 more realistic 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 dataset. 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.

B. Associated Problems

1) 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 socio-cultural, 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 social media framework. Facebook allows users to “like” pictures, 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.

A user study to determine factors that would prevent people from including a picture in their albums was reported in [65]. 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 (IVI) to pictures based on their technical and aesthetic qualities and social relevance has been proposed in [47]. 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 [47].

Discussion: While a personal or situational appeal or value would be of greater interest to a non-specialist user, generic models for appeal may be even more short-lived than for aesthetics. In order 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 [27].

2) 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 [7][8] [21][39][40][64]. 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 [32].

Brushstrokes provide reliable modeling information for certain types of paintings that do not have colors. In [45], 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 [6]. 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 (and in turn influence of artists on one another). A valuable problem to the commercial art community is to model and predict a common-man's perception and appreciation of art as opposed to that of art connoisseurs [44] .

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 [98]. 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 [86] wherein training is performed using a well-known image dataset in psychology while the approach is demonstrated on 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. 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 inputs 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 inputs 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 [32].

3) 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. Understanding beauty has been an important discipline in experimental psychology [79]. 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 [66][74][88]. Therefore, it seems that computer vision research on asymmetry in faces, such as [46], 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 [18]. The cited work uses advanced 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 [17].

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 beauty contests also assess the complete personality of participants and do not judge merely by visual aspects.

4) 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 [80]. Human factors such as mentioned above largely provide a way to rerank images or search among equals for matches closer to the heart of a user. In [4], an image filtering system is described that uses the Kansei methodology to associate low-level image features with human feelings and impressions. Another work [22] attempts to model the target image within the mind of a user using relevance feedback to learn a distribution over the image database. In a recent work, the attractiveness of images is used to enhance the performance of Web image search engine (in terms of the online ranking, interactive re-ranking, and offline index selection) in [28]. Along similar lines, [63] integrates semantic, aesthetic, and affective features to achieve significant improvement for the task of scene recognition on various diverse and large-scale datasets.

Discussion: 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. One potential research direction is to assess the tradeoff between personalization of results and speed of retrieval.

IV. COMPUTATIONAL FRAMEWORK

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 representation* and *modeling and 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 [69]. In an analogous fashion, the technical challenge in automatic inference of aesthetics is defined in [16] 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.

A. 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 [14]. 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 [2][26][59]. Conceiving meaningful visual properties that may have correlation with perceived aesthetics or an emotion is itself a challenging problem. In literature, we notice a spectrum from very generic color, texture, and shape features to specifically designed feature descriptors to model 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 attempt to 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 (Fig. 4). In [13], 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 (Fig. 4). In [13], wavelets have been used to detect a picture with a low depth of field. 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 narrow DOF are not considered pleasing; instead, photographers prefer to have the foreground, middle ground, and background all in focus).

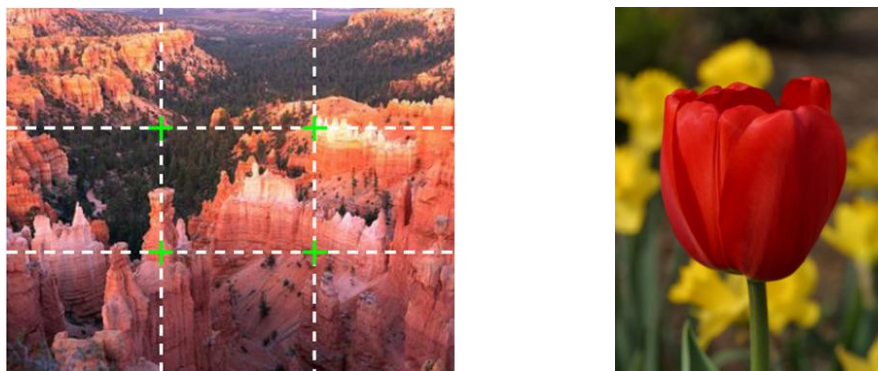


Figure 4: Left: The Rule of Thirds in photography; Right: A low depth-of-field picture.

A mix of global and local features has been used in [44] to model the aesthetics problem for paintings. Feature selection is based on the belief that people use a top-down approach to appreciate art. 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 [34]. 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 depth of field, and (ii) a surrealism created by the proper choice of camera parameters and appropriate lighting conditions.

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 also been shown to be very helpful. A computational visual attention model using a face-sensitive saliency map is proposed in [73]. 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 [85] where each pixel is assigned a mass based on its luma and chroma values (YCbCr space) and pixels exert a gravity-like mutual force.

Some recent papers focus on enhancement of images or suggestion of ideal composition based on aesthetically learned rules [5][11]. Two distinct recomposition techniques based on key aesthetic principles (“rule of thirds” and “golden ratio”) have been proposed in [5]. 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). Another interesting work [11] 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. Contextual modeling involves learning a spatial Gaussian mixture model for pairwise visual words. A recent work [51] explores the role of content in image aesthetics by designing specific visual features for different categories (e.g. landscape, plant, animal, night, human, static, and architecture). The work focuses on detecting and extracting local features from the most attractive image region (from among region of focus, vertical standing objects, or human faces).

Several recent papers have emphasized the usability of generic descriptors constructed by local features for image aesthetics. Along this line bag-of-visual-words and Fisher vectors (that encode more local information) have been explored to improve the accuracy of image aesthetics assessment in [53]. Gradient information is extracted through SIFT and color features and significant improvements (over previous works) have been reported. The influence of the color harmony of photos on the aesthetic quality has been investigated in

[55]. By representing photos as a collection of local regions, the work models the color harmony (as predictor of aesthetic quality) of photos through bags-of-color-patterns. Patch wise bag-of-aesthetics-preserving features that encode contrast information are explored in [72]. O'Donovan et al. model the quality of color themes that refer to a five-color palette by learning from a large-scale dataset with a regression method in [19].

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 [86], the authors divert from the common codebook approach to a methodology where similarity to all vocabulary elements is preserved for emotion category modeling. In [6], low-level local visual features including SIFT 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 [4] using 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 [52]. Among other features, [52] adopts the standardized Pleasure-Arousal-Dominance transform color space, composition features such as low-depth-of-field indicators and rule of thirds (which have been found to be useful for aesthetics), and proportion of skin pixels in images. In [61], 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 also explored in [1] to understand and predict topical relevance. The work models neurological signals and facial expressions of users looking at images as implicit relevance feedback. In order to

classify emotions, [1] employs a 3-D wire-frame model of faces and tracks presence and degrees of changes in different facial regions. Similarly, [78] also employs face tracking to extract facial motion features for emotion classification.

A recent work, [48] explores the relationship between shape characteristics (such as roundness, angularity, simplicity, and complexity) and emotions. Shape features constitute line segments, continuous lines, angles, and curves, to reflect such characteristics. In an interesting diversion, inferring aroused emotions from images in social networks has been studied in [31]. The work represents the emotion by 16 discrete categories that cover the affective space. Color features (e.g., saturation, brightness, and HSV) and social features (e.g., uploading time and user ID) were extracted as image descriptors.

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 [17][20] as well as non-geometric ones (such as hair color and skin smoothness) [20].

B. Modeling and Learning

Aesthetics and Emotion modeling literature reports use of both discriminative learning methods such as SVM and CART [13] [44][47][86] and generative learning techniques such as naïve Bayes, Bayesian networks, and Gaussian mixture models [52] [49][78][11]. While two-class or multi-class classification paradigm seems to be the norm, support vector and kernel regression methods have also been explored [5] [17]. An adapted regression approach to map visual features extracted from photos to a distribution has been presented in [83]. A dimensional approach to represent emotions (to capture correlations between emotional words) has been explored in [48]. [31] presents a partially labeled factor graph model to infer the emotions aroused from images within a social network setting. A bilayer sparse representation is proposed to encode similarities among global images, local regions, and the

regions' co-occurrence property in [43]. The proposed context-aware classification model with the bilayer sparse representation shows a higher accuracy in predicting categorized emotions on the IAPS dataset. In conclusion, we can state that while learning lies at the heart of every computational inference problem that we consider here, choices of the modeling and learning strategies vary with the nature of the task and features.

V. DATA RESOURCES

A. *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 (EEG), electrooculogram, heart rhythm, pupillary reactions, and more recently, neural activity in various brain areas obtained using functional magnetic resonance imaging (fMRI) [37][18]. Recording eye movements is also a valuable technique that helps detect where the viewers are looking when evaluating aesthetic attributes of art compositions [56].

Certain efforts have resulted in the creation of a specialized database for emotion studies known as the International Affective Picture Systems (IAPS) database (Fig. 5) [42]. 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 [86].



Figure 5: (top) Pictures of Yosemite National Park from Terragallaria.com, (bottom) Example images from IAPS (The International Affective Picture System) dataset. Images with a more positive affect from left to right, and higher arousal from bottom to top.

B. 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 the useful information on the Web [30]. 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 [94] 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, 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 [96] is a platform for photography enthusiasts to share and have their pictures peer-rated on a 1–7 scale of aesthetics. The photography community also provides discussion forums, reviews on photos and photography products, and galleries for members and casual surfers.

DPChallenge [93] 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 1–10 scale, determines the contest winners.

Terragalleria [97] showcases travel photography of Quang-Tuan Luong (a scientist and a photographer), and is one of the finest resources for US national park photography on the Web (Fig. 5). All photographs here have been taken by one person (unlike Photo.Net), but multiple users have rated them on overall quality on a 1–10 scale.

ALIPR [92] is a Web-based image search and tagging system that also allows users to rate photographs along 10 different emotional categories such as surprising, amusing, pleasing, exciting, and adorable.

Besides this, certain research efforts have created their own collections of data from the above sources notably (i) a manually labeled dataset with over 17,000 photos covering seven semantic categories [51], and (ii) AVA dataset to facilitate aesthetics visual analysis [54] consisting of about 250,000 images from DPChallenge.

C. 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 datasets. These can be used to compare competing algorithms in the future. Here we present plots of features of the datasets, in particular the nature of user ratings received in each case (not necessarily comparable across the datasets).

Fig. 6 shows the distribution of mean aesthetics. We begin with a section called Features Plots of Aesthetics Ratings in which we describe the nature of the plots. In the following section, called Analysis of Feature Plots, we conduct a thorough analysis of each figure, breaking it up for each data source/quality score received by each photo. Fig. 7 shows the distribution of the number of ratings each photo received. In Fig. 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 Fig. 9, we plot the distribution of the fraction of ratings received by each photo within ± 0.5

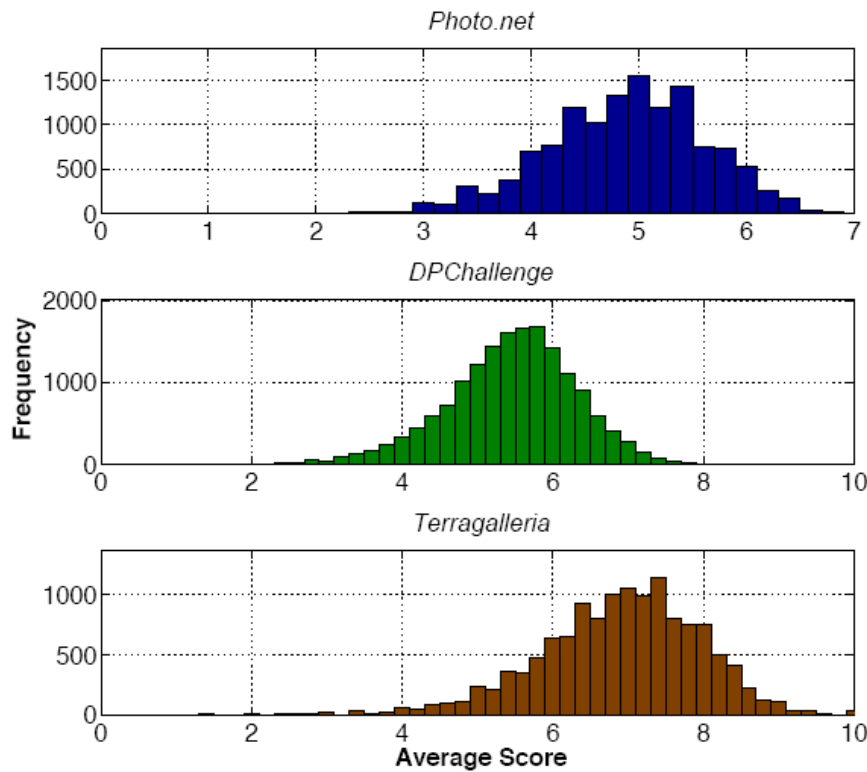


Figure 6: Distributions of average aesthetics scores from three different data collections.

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 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 Fig. 10, we plot the distribution of emotions votes in the dataset 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 Figs. 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 than 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 non-blind 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 Fig. 6, we see that for each dataset, 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 5, which is a full point above the mid-point 4. 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 Fig. 6 is that the distribution is smoother for DPChallenge than for the other two. This may simply be because this dataset has the largest

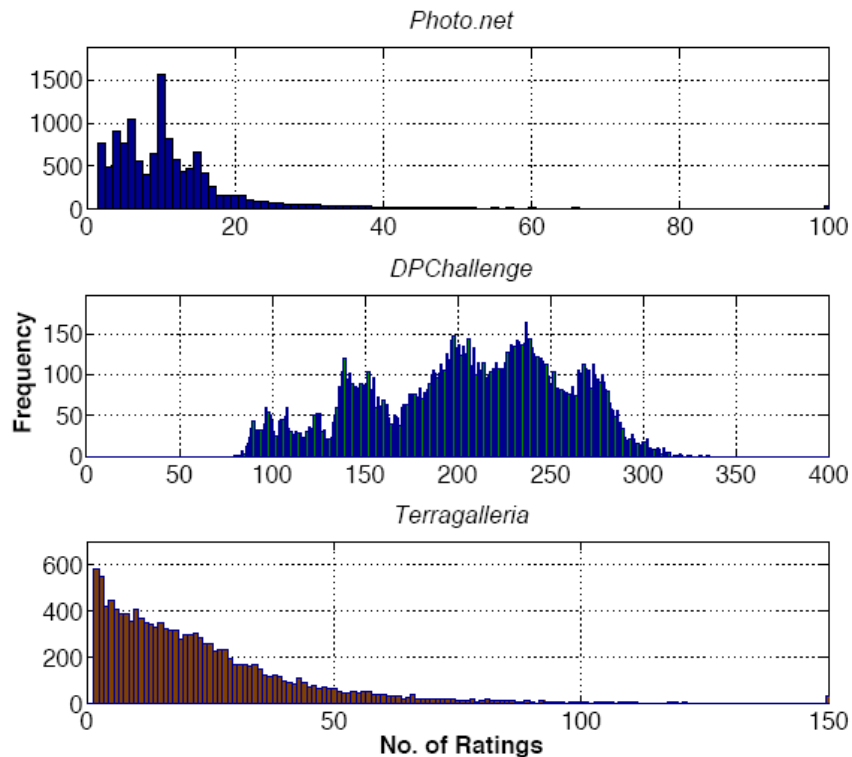


Figure 7: Distributions of number of ratings from three different data collections.

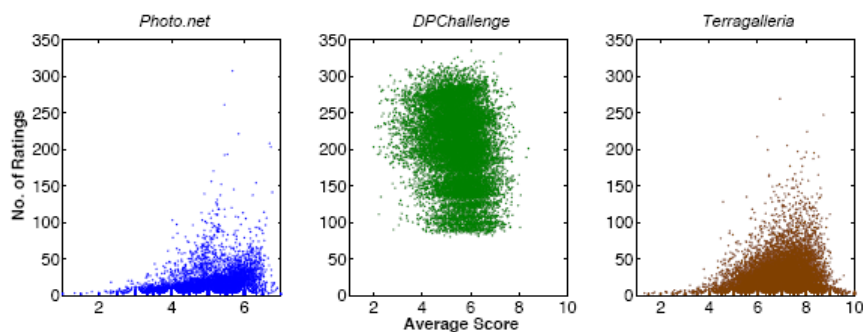


Figure 8: Correlation plot of (avg. score, no. of ratings) pairs.

sample size. In Fig. 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

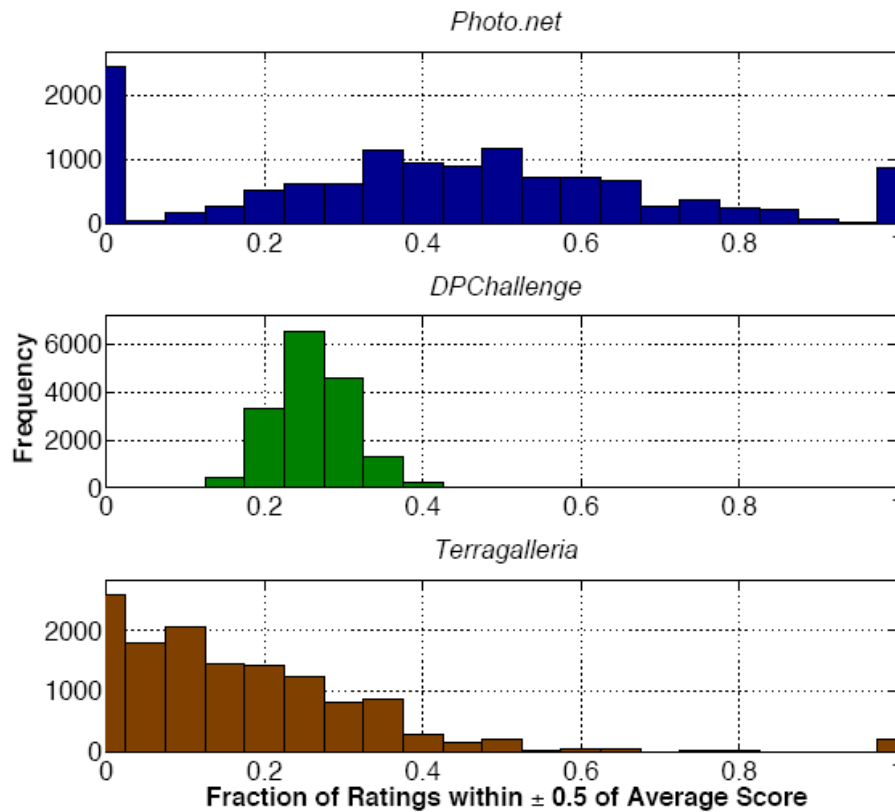


Figure 9: Distribution of the level of consensus among ratings.

almost resembles a mixture of Gaussians with means at well-spaced locations. It is unclear to the authors as to the nature of the social phenomenon on DPChallenge.com that these peaks might be associated with. Photos on Photo.net are much rarer, mainly because the process is non-competitive, 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 Fig. 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

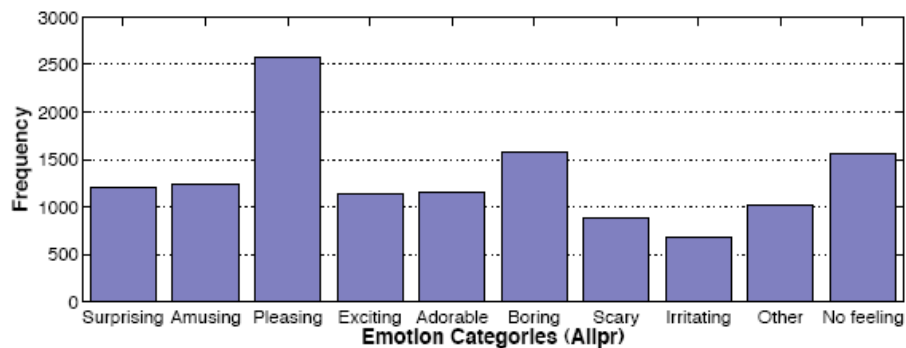


Figure 10: Distribution of emotion votes given to images (ALIPR).

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 in Websites such as IMDB. Associated with the issue that people tend to explicitly rate mainly things they like is the fact that the Websites 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 Fig. 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 Fig. 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 that graph). Finally, in Fig. 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.”

VI. FUTURE RESEARCH DIRECTIONS

A. Understanding Socio-cultural, Personal, and Psychology Induced 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. Additional personalization can be achieved by understanding tastes of individuals which will however require significant amount of personalized data for model building.

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 [9]. Studies have also shown that art preferences and art judgment can vary significantly across expert and non-expert subjects [29]. Artists and experienced art viewers tend to prefer artworks that are challenging and emotionally provocative [82], which is in contrast to the majority of people who prefer art that makes them happy and feel relaxed [84]. The results reported in [2], [10], and [25] demonstrate that such differences are significant and can be explained on the basis of common mechanisms as suggested by Berlyne in [3].

B. Understanding and Modeling Context

Context plays an important role in semantic image understanding [50]. 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) [35][36], and social context [27] [68][90] (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.

C. 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 [91], an attempt in this direction. ACQUINE (Aesthetic Quality Inference Engine) is a machine-learning-based online system that showcases computer-based prediction of aesthetic quality for color natural photographic pictures (Fig. 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. Another recently developed system OSCAR (On-site Composition and Aesthetics Feedback) aims at helping photographers to generate high-quality photos [87]. OSCAR provides on-site analyses of photos in terms of the composition and aesthetics quality and generates feedback through high-quality examples.

We envision a future where consumer cameras and smartphones are equipped with an automated personal assistant that can provide aesthetics judgment so that only the highest quality photos are taken and stored. Such a module can be a post-photography filter or a real-time filter (such a real-time aesthetics-meter). A recent effort in this direction is the Nadia

camera that uses ACQUINE to offer a real-time aesthetics score [95]. Real-time photography feedback is not a stranger today's photographers (face detection, smile detection etc.). Hence the dream of aesthetics feedback in cameras may not be that distant.

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 datasets 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.

VII. ACKNOWLEDGEMENTS

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 is done while working at the National Science Foundation. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Foundation.

REFERENCES

- [1] I. Arapakis, I. Konstas, and J. M. Jose, "Using Facial Expressions and Peripheral Physiological Signals as Implicit Indicators of Topical Relevance", *Proc. ACM Multimedia*, 2009.

- [2] O. Axelsson, "Towards a Psychology of Photography: Dimensions Underlying Aesthetic Appeal of Photographs," *Perceptual and Motor Skills*, 105:411–434, 2007.
- [3] D.E. Berlyne, "Aesthetics and Psychobiology," *New York: Appleton-Century-Crofts*, 1971.
- [4] N. Bianchi-Berthouze, "K-dime: An Affective Image Filtering System," *IEEE Multimedia*, 10(3):103–106, 2003.
- [5] S. Bhattacharya, R. Sukthankar, M. Shah, "A Framework for Photo-Quality Assessment and Enhancement Based on Visual Aesthetics", *Proc. ACM Multimedia*, 2010.
- [6] M. Bressan, C. Cifarelli, and F. Perronnin, "An Analysis of the Relationship between Painters Based on Their Work," *Proc. IEEE ICIP*, 2008.
- [7] I.E. Berezhnuy, E.O. Postma, and H.J. van den Herik, "Computerized Visual Analysis of Paintings," *Proc. 16th Int. Conf. Assoc. for History and Computing*, 28–32, 2005.
- [8] I.E. Berezhnuy, E.O. Postma, and H.J. van den Herik, "Computer Analysis of Van Gogh's Complementary Colors," *Pattern Recognition Letters*, 28(6):703–709, 2007.
- [9] G.H. Bower, "Mood and Memory," *American Psychologist*, 36:129–148, 1981.
- [10] C. Cerosaletti and A. Loui, "Measuring the Perceived Aesthetic Quality of Photographic Images," *Proc. 1st International Workshop on Quality of Multimedia Experience*, 2009.
- [11] B. Cheng, B. Ni, S. Yan, and Q. Tian, "Learning to Photograph," *Proc. ACM Multimedia*, 2010.
- [12] S. Daly, "The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity," *Digital Image and Human Vision*, 179–206, 1993.
- [13] R. Datta, D. Joshi, J. Li, and J.Z. Wang, "Studying Aesthetics in Photographic Images Using a Computational Approach," *Proc. ECCV*, 2006.
- [14] R. Data, D. Joshi, J. Li, and J.Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," *ACM Computing Surveys*, 40(2), 5:1–60, 2008.

- [15] R. Datta, J. Li, and J.Z. Wang, “Learning the Consensus on Visual Quality for Next Generation Image Management,” *Proc. ACM Multimedia*, 2007.
- [16] R. Datta, J. Li, and J.Z. Wang, “Algorithmic Inferencing of Aesthetics and Emotion in Natural Images: An Exposition,” *Proc. ICIP*, 2008.
- [17] B.C. Davis and S. Lazebnik, “Analysis of Human Attractiveness using Manifold Kernel Regression,” *Proc. ICIP*, 2008.
- [18] 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*, 41:147–155, 2003.
- [19] P. O.’Donovan, A. Agarwala, and A. Hertzmann, “Color Compatibility from Large Datasets,” *ACM Transactions on Graphics*, 30,4, 63:1, 63:12, 2011.
- [20] Y. Eisenthal, G. Dror, and E. Ruppín, “Facial Attractiveness: Beauty and the Machine,” *Neural Computation*, 18(1):119–142, 2006.
- [21] C.M. Falco, “Computer Vision and Art,” *IEEE Multimedia*, 14(2):8–11, 2007.
- [22] Y. Fang, D. Geman, and N. Boujemaa, “An Interactive System for Mental Face Retrieval,” *Proc. ACM SIGMM Int. Workshop on Multimedia Information Retrieval*, 2005.
- [23] G.T. Fechner, “Zur experimentalen Ästhetik (On Experimental Aesthetics),” *Abhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften*, 9:555–635, 1871.
- [24] G.T. Fechner, “Vorschule der Aesthetik,” *Breitkopf und Härtel, Leipzig*, 2 volumes, 1876.
- [25] E.A. Fedorovskaya, C. Neustaedter, and W. Hao, “Image Harmony for Consumer Images,” *Proc. ICIP*, 2008.
- [26] M. Freeman, “The Photographer’s Eye,” *Focal Press*, 2007.

- [27] A. Gallagher and T. Chen, “Using Context to Recognize People in Consumer Images,” *IPSJ Transactions on Computer Vision and Applications*, 2009.
- [28] B. Geng, L.J. Yang, C. Xu, X.-S. Hua, and S.P. Li, “The Role of Attractiveness in Web Image Search,” *Proc. ACM Multimedia*, 2011.
- [29] P. Hekkert, “Beauty in the Eye of Expert and Non-Expert Beholders: A Study in the Appraisal of Art,” *The American J. of Psychology*, 109(3):389–407, 1997.
- [30] J. Howe, “The Rise of Crowdsourcing,” *Wired Magazine*, 2006.
- [31] J. Jia, S. Wu, X.H. Wang, P.Y. Hu, L.H. Cai, and J. Tang, “Can We Infer Van Gogh’s Mood? Learning to Infer Affects from Images in Social Networks,” *Proc. ACM Multimedia*, 2012.
- [32] 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 Magazine, Special Issue on Visual Cultural Heritage*, 25(4):37–48, 2008.
- [33] H. Kawabata and S. Zeki, “Neural Correlates of Beauty,” *J. of Neurophysiology*, 91:1699–1705, 2004.
- [34] Y. Ke, X. Tang, and F. Jing, “The Design of High-Level Features for Photo Quality Assessment,” *Proc. CVPR*, 2006.
- [35] L. Kennedy, M. Naaman, “How Flickr Helps Us Make Sense of the World: Context and Content in Community-contributed Media Collections,” *Proc. ACM Multimedia*, 2007.
- [36] L. Kennedy and M. Naaman, “Generating Diverse and Representative Image Search Results for Landmarks,” *Proc. of the 17th Int. Conf. on World Wide Web*, 2008.
- [37] U. Kirk, M. Skov, O. Hulme, M.S. Christensen, and S. Zeki, “Modulation of Aesthetic Value by Semantic Context: An fMRI Study,” *Neuroimage*, 44:1125–1132, 2009.
- [38] K. Koffka, “Gestalt Psychology,” *Jovanovic: Harcourt Brace*, 1935.

- [39] S. Kroner and A. Lattner, "Authentication of Free Hand Drawings by Pattern Recognition Methods," *Proc. IEEE ICPR*, 1998.
- [40] A. Kushki, P. Androustos, K. Plataniotis, and A. Venetsanopoulos, "Retrieval of Images from Artistic Repositories using a Decision Fusion Framework," *IEEE Trans. on Image Processing*, 13(3):277–292, 2004.
- [41] P.J. Lang, M.K. Greenwald, M.M. Bradley, and A.O. Hamm, "Looking at Pictures: Affective, Facial, Visceral, and Behavioral Reactions," *Psychophysiology*, 30:261–273, 1993.
- [42] 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*, 1997.
- [43] B. Li, W.H. Xiong, W.M. Hu, and X.M. Ding, "Context-aware Affective Images Classification based on Bilayer Sparse Representation", *Proc. ACM Multimedia*, 2012.
- [44] C.C. Li and T. Chen, "Aesthetic Visual Quality Assessment of Paintings," *IEEE Journal of Selected Topics in Signal Processing*, 3(2):236–252, 2009.
- [45] J. Li and J.Z. Wang, "Studying Digital Imagery of Ancient Paintings by Mixtures of Stochastic Models," *IEEE Trans. on Image Processing*, 13(3):340–353, 2004.
- [46] Y. Liu, K.L. Schmidt, J.F. Cohn, and S. Mitra, "Facial Asymmetry Quantification for Expression Invariant Human Identification," *Proc. CVPR*, 2003.
- [47] A. Louis, M.D. Wood, A. Scalise, and J. Birkelund, "Multidimensional Image Value Assessment and Rating for Automated Albuming and Retrieval," *Proc. ICIP*, 2008.
- [48] X. Lu, P. Suryanarayan, R.B. Adams, J. Li, M.G. Newman, and J.Z. Wang, "On Shape and Computability of Emotions", *Proc. ACM Multimedia*, 2012.
- [49] J. Luo, A. Savakis, S. Etz, and A. Singhal, "On the Application of Bayes Networks to Semantic Understanding of Consumer Photographs", *Proc. ICIP*, 2000.

- [50] J. Luo, M. Boutell, and C. Brown, "Exploiting Context for Semantic Scene Content Understanding," *IEEE Signal Processing Magazine, Special Issue on Semantic Retrieval of Multimedia*, 23(2):101–114, 2006.
- [51] W. Luo, X.G. Wang, and X.O. Tang, "Content-based Photo Quality Assessment", *Proc. ICCV*, 2011.
- [52] J. Machajdik and A. Hanbury, "Affective Image Classification using Features Inspired by Psychology and Art Theory", *Proc. ACM Multimedia*, 2010.
- [53] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka, "Assessing the Aesthetic Quality of Photographs using Generic Image Descriptors", *Proc. ICCV*, 2011.
- [54] N. Murray, L. Marchesotti, and F. Perronnin, "AVA: A Large-Scale Database for Aesthetic Visual Analysis", *Proc. CVPR*, 2012.
- [55] M. Nishiyama, T. Okabe, I. Sato, and Y. Sato, "Aesthetic Quality Classification of Photographs based on Color Harmony", *Proc. CVPR*, 2011.
- [56] C.F. Nodien, P.J. Locher, and E.A. Krupinski, "The Role of Formal Art Training on Perception and Aesthetic Judgment of Art Compositions," *Leonardo*, 26(3):219–227, 1993.
- [57] S.E. Palmer, "Aesthetic Science," *Proc. IS&T/SPIE Electronic Imaging Conference*, 2009.
- [58] D.I. Perrett, K.A. May, and S. Yoshikawa, "Facial Shape and Judgments of Female Attractiveness," *Nature*, 368:239–242, 1994.
- [59] G. Peters, "Aesthetic Primitives of Images for Visualization," *Proc. IEEE Int. Conf. Information Visualization*, 2007.
- [60] V.S. Ramachandran and W. Hirstein, "Science of Art: A Neurological Theory of Aesthetic Experience," *J. of Consciousness Studies*, 6:15–51, 1999.

- [61] S. Ramanathan, H. Katti, R. Huang, T.-S. Chua, and M. Kankanhalli, “Automated Localization of Affective Objects and Actions in Images via Caption Text-cum-Eye Gaze Analysis,” *Proc. ACM Multimedia*, 2009.
- [62] R.N. Reber, N. Schwartz, and P. Winkielman, “Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver’s Processing Experience?,” *Personality and Social Psychology Review*, 8(4):364–382, 2004.
- [63] M. Redi, and B. Merialdo, “Enhancing Semantic Features with Compositional Analysis for Scene Recognition”, *Proc. ECCV*, 2012.
- [64] D. Rockmore, S. Lyu, and H. Farid, “A Digital Technique for Authentication in the Visual Arts,” *International Foundation for Art Research*, 8(2):12–23, 2006.
- [65] A. Savakis, S. Etz, and A. Loui, “Evaluation of Image Appeal in Consumer Photography,” *Proc. SPIE Human Vision and Electronic Imaging*, 2000.
- [66] J.E. Scheib, S.W. Gangestad, and R. Thornhill, “Facial Attractiveness, Symmetry, and Cues of Good Genes,” *Proc. Royal Society of London*, 266:1913–1917, 1999.
- [67] H.R. Sheikh, A.C. Bovik, and L. Cormack, “No-reference Quality Assessment using Natural Scene Statistics: JPEG2000,” *IEEE Trans. Image Processing*, 14(11):1918–1927, 2005.
- [68] B. Shevade and H. Sundaram, “Modeling Personal and Social Network Context for Event Annotation in Images,” *Proc. Joint Conf. on Digital Libraries*, 2007.
- [69] 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. on Pattern Analysis and Machine Intelligence*, 22(12), 2000.
- [70] R.L. Solso, “The Psychology of Art and the Evolution of the Conscious Brain,” *MIT Press*, 2003.

- [71] D. Stork, "Computer Vision and Computer Graphics Analysis of Paintings and Drawings: An Introduction to the Literature," *Proc. International Conference on Computer Analysis of Images and Patterns*, LNCS 5702, 9–24, 2009.
- [72] H.-H. Su, T.-W. Chen, C.-C. Kao, W. H. Hsu, and S.-Y. Chien, "Scenic Photo Quality Assessment with Bag of Aesthetics-Preserving Features", *Proc. ACM Multimedia*, 2011.
- [73] X. Sun and H. Yao, "Photo Assessment Based on Computational Visual Attention Model", *Proc. ACM Multimedia*, 2009.
- [74] J.P. Swaddle and I.C. Cuthill, "Asymmetry and Human Facial Attractiveness: Symmetry May Not Always be Beautiful," *Proc. Biological Sciences*, 1995.
- [75] R. Taylor, A.P. Micolich, and D. Jones. "Fractal Analysis of Pollock's Drip Paintings," *Nature*, 399(June 3):422(1999), 1999.
- [76] R. Taylor, "Pollock, Mondrian and the Nature: Recent Scientific Investigations," *Chaos and Complexity Letters*, 2004.
- [77] R. Valenti, N. Sebe, and T. Gevers, "Facial Expression Recognition: A Fully Integrated Approach," *Proc. Intl. Workshop on Visual and Multimedia Digital Libraries*, 2007.
- [78] R. Valenti, A. Jaimes, and N. Sebe, "Sonify Your Face: Facial Expressions for Sound Generation", *Proc. ACM Multimedia*, 2010.
- [79] C.W. Valentine, "The Experimental Psychology of Beauty," *Menthuen and Co. Publishers*, 1962.
- [80] W. Wang and Q. He, "A Survey on Emotional Semantic Image Retrieval," *Proc. ICIP*, 2008.
- [81] A.B. Watson, "Toward a Perceptual Video Quality Metric," *Proc. SPIE*, 3299:139–147, 1998.
- [82] A.S. Winston and G.C. Cupchik, "The Evaluation of High Art and Popular Art by Naive and Experienced Viewers," *Visual Arts Research*, 18(1):1–14, 1992.

- [83] O. Wu, W.M. Hu, and J. Gao, "Learning to Predict the Perceived Visual Quality of Photos", *Proc. ICCV*, 2011.
- [84] J. Ed. Wypijewski, *Painting by the Numbers: Komar and Melamid's Scientific Guide to Art*, New York, *Farrar, Straus & Giroux*, 1997.
- [85] Y. Yang, M. Song, N. Li, J. Bu, and C. Chen, "Visual Attention Analysis by Pseudo Gravitational Field", *Proc. ACM Multimedia*, 2010.
- [86] V. Yanulevskaya, J.C. van Gemert, K. Roth, A. K. Herbold, N. Sebe, and J.M. Geusebroek, "Emotional Valence Categorization using Holistic Image Features," *Proc. ICIP*, 2008.
- [87] L. Yao, P. Suryanarayan, M. Qiao, J.Z. Wang, J. Li, "OSCAR: On-site Composition and Aesthetics Feedback through Exemplars for Photographers", *Int. Journal of Computer Vision*, 96, 3, 353-383.
- [88] D.W. Zaidel and J.A. Cohen, "The Face, Beauty, and Symmetry: Perceiving Asymmetry in Beautiful Faces," *Int. J. of Neuroscience*, 115:1165–1173, 2005.
- [89] S. Zeki, "Inner Vision: An Exploration of Art and the Brain," *Oxford University Press*, 1999.
- [90] A. Zunjarwad, H. Sundaram, "Contextual Wisdom: Social Relations and Correlations for Multimedia Event Annotation", *Proc. ACM Multimedia*, 2007.
- [91] ACQUINE, <http://acquine.alipr.com>.
- [92] ALIPR, <http://alipr.com>.
- [93] DPChallenge, <http://www.dpchallenge.com>.
- [94] Flickr, <http://www.flickr.com>.
- [95] Nadia Camera, <http://www.wired.com/gadgetlab/2010/07/nadia-camera-offers-opinion-of-your-terrible-photos/>.
- [96] Photo.net, <http://photo.net>.

[97] Terragalleria, <http://www.terrageria.com>.

[98] USA Today, http://www.usatoday.com/tech/news/techinnovations/2006-12-18-computer-feelings_x.htm