Automated Aesthetic Analysis of Photographic Images

Tunç Ozan Aydın, Aljoscha Smolic, and Markus Gross

Abstract—We present a perceptually calibrated system for automatic aesthetic evaluation of photographic images. Our work builds upon the concepts of no-reference image quality assessment, with the main difference being our focus on rating image aesthetic attributes rather than detecting image distortions. In contrast to the recent attempts on the highly subjective aesthetic judgment problems such as binary aesthetic classification and the prediction of an image's overall aesthetics rating, our method aims on providing a reliable objective basis of comparison between aesthetic properties of different photographs. To that end our system computes perceptually calibrated ratings for a set of fundamental and meaningful aesthetic attributes, that together form an "aesthetic signature" of an image. We show that aesthetic signatures can still be used to improve upon the current state-of-the-art in automatic aesthetic judgment, but also enable interesting new photo editing applications such as automated aesthetic analysis, HDR tone mapping evaluation, and providing aesthetic feedback during multi-scale contrast manipulation.

Index Terms—Image aesthetics assessment, aesthetic signature, aesthetic attributes.

1 Introduction

Recent advancements in image acquisition and visual computing made technology cheaper and easily available, consequently putting more power into the hands of an average user. High quality cameras, either standalone or integrated into mobile devices, as well as advanced image editing tools are more commonly used than ever. From the user's point of view, these new technologies create the expectation of more appealing images. But obtaining appealing results requires not only advanced tools, but also the knowledge and execution of basic aesthetic principles during acquisition and editing. The problem is that the average user does not always have the necessary training and experience, nor the interest in acquiring them. Thus, modeling aesthetic principles and building systems that give automatic aesthetic feedback is a research area with high practical relevance.

The interest in obtaining aesthetically pleasing results with minimal effort is evident from the fact that simple and effective photo editing tools like Instagram are very popular among casual photographers. Similarly, Fujifilm's Image Intelligence™ framework that utilizes multiple systems (such as light source recognition, face detection, etc.) to improve image aesthetics, Sony's Party-shot™ technology where a rotating platform adjusts the camera for best photographic composition, and the Smile Shutter™ where the camera releases the shutter when people smile, are all examples for integration of models of basic photographic principles with the current imaging technologies. These developments in the industry are also paralleled by the research community with the recently increasing amount of publications on two-fold classification of image sets into aesthetically appealing or not, and automatic aesthetic judgment by predicting an overall aesthetic rating (discussed in detail in Section 2).

While automatic aesthetic judgment is useful for many practical purposes, such judgments the in form of a yes/no answer, or a percentage score do not explain *why* the evaluated image is aesthetically pleasing or not. This is because when designing such systems, understandably the image features are selected based on their classification performance of overall aesthetics, but not necessarily on how well they correlate with the aesthetic attribute they claim to evaluate. As an example, it is often not discussed if a "clarity" feature actually corresponds to what people consider as the photographic clarity rule, or is some abstract heuristic that happened to result in accurate classification. While this approach is perfectly fine for predicting a single-dimensional outcome, a multi-dimensional aesthetic analysis based on ratings of meaningful aesthetic attributes requires a different approach and poses additional challenges.

The first challenge is finding a set of image attributes that are simple enough to be expressed as computer programs, but at the same time are closely related to some fundamental photographic attributes. Once these attributes are defined, another challenge is designing and executing a subjective study through which one can reliably determine ground truth attribute ratings on a set of real world images. Once the subjective data is obtained, the final challenge is the design, implementation and calibration of metrics that predict a rating for each aesthetic attribute.

- *T. O. Aydın and A. Smolic are with Disney Research Zürich. E-mail: tunc@disneyresearch.com*
- *Markus Gross is with ETH Zürich and Disney Research Zürich*

Fig. 1. What are the aesthetic differences between the original and edited images? Given a single image our method automatically generates a calibrated "aesthetic signature" as a compact, objective representation of its aesthetic properties. A comparison of the aesthetic signatures reveals that the editing greatly enhanced tone and depth of the under-exposed original image at the cost of slight decreases in sharpness and clarity. Images courtesy of Wojciech Jarosz.

In this work we tackle these challenges and present a system that computes an *aesthetic signature* from a single image, that comprises calibrated ratings of meaningful aesthetic attributes and provides an objective basis for aesthetic evaluation (Figure 1). In the paper we introduce these aesthetic attributes (Section 3), discuss the experimental procedure through which we obtain subjective ratings for each aesthetic attribute (Section 4), and propose metrics that predict aesthetic attribute ratings and calibrate them using subjective data (Section 5). We also present exemplary applications of our system to automated aesthetic analysis, HDR tone mapping evaluation, and multi-scale contrast editing (Section 6).

In the next section we review the previous work on automated image aesthetics, image quality assessment and subjective evaluation of high level visual attributes. We also briefly discuss the general limits and scope of computational aesthetic judgment.

2 Related Work

There are multiple sources on the basic guidelines of photography (refer to Freelan's work [1] as an example). These publications often describe a set of photographic principles that should be taken into consideration for shooting aesthetically pleasing photographs. At a high level, the task of the photographer can be seen as evaluating the shot in terms of these photographic principles and seeking the optimal balance between different aesthetic attributes that leads to an aesthetically pleasing result.

Computational aesthetic judgment methods follow a workflow similar to the photographer's. Aesthetic judgment has often been approached as a learning problem on image features obtained from a large set of images (see Savakis et al. [2], Datta et al. [3] and Joshi et al. [4] for an overview), where the task is a binary classification between aesthetically pleasing and not pleasing images. Datta, et al. [5] proposed a linear SVM classifier that uses 15 image features selected among the initially proposed 56 features based on classification performance. A similar approach for video has been presented by Moorthy et al. [6]. Ke et al. [7] proposed using image features based on common-sense photography and utilize a naïve Bayes classifier based on their observation that the interactions between the aesthetic attributes are not linear. The two-fold classification accuracy of all these methods on subjective data is in the 70% range. More recent work along these lines evaluated the use of generic image descriptors for aesthetic quality assessment [8]. Luo and Tang [9] reported a significant improvement in accuracy by extracting a rectangular image window that contains most of the high frequency details, and formulating features that take into account this two-fold segmentation. Other work in this area includes [10], [11], [12], [13]. Similarly, segmentation using a saliency map [14] and face detection [15] has been explored in the context of automated image aesthetics. Unlike the binary classification methods, *acquine* is a popular online aesthetic rating engine [16] that predicts an overall aesthetic percentage rating. Recent work has also been focused on more specific sub-problems such as photographic composition [17], [18], [19], [20], view recommendation [21], color compatibility [22], [23], [24], and candid portrait selection from videos [25], as well as the use of more specialized features like sky-illumination attributes and object types in the scene [26]. Finally, a large data set with associated meta-data has been published to facilitate further image aesthetics research. [27] Our work takes inspiration from the great body of previous work in this area, with the main difference being our emphasis on the novel *aesthetic signature* concept.

Image quality assessment methods seek to estimate "image quality" without requiring user involvement. Given a test image with some imperfections, quality is either defined as the fidelity to a reference image [28], [29], or by the absence of certain types of distortions such as compression, ringing [30], blur [31], and banding [32]. The latter, no-reference type of quality assessment is significantly more challenging

because such quality metrics do not utilize a reference image, but instead rely on their internal model of distortions. There have been also some attempts on building more general no-reference metrics by combining the individual contribution of image features an distortions [33], as well as utilizing natural image statistics [34]. An interesting recent work on image completion combines concepts from image aesthetics and quality prediction [35]. At a conceptual level, our method is influenced by such generalized no-reference quality assessment methods. However, our work is fundamentally different in that our metrics predict the magnitude of a set of aesthetic attributes instead of predicting the visibility of distortions. Moreover, as its outcome our method provides a basis for aesthetic analysis rather than assessing image quality.

Subjective evaluation of visual attributes has been performed through psychophysical experiments with the goal of determining a mapping from objectively computable values to perceptually meaningful units. A classic example is Whittle's luminance difference discrimination experiment [36] that reveals the non-linear perception of luminance contrast. More recently, similar experimental methods have been used to derive models of "visual equivalence" of objects with different material properties, geometry and illumination [37], the discriminability of aggregates of objects [38], the effects of global illumination approximations on material appearance [39] among others. Our experimental method is analogous to this line of research, in that we investigate the perception of aesthetic attributes and seek to design and calibrate metrics whose predictions match subjective ground truth ratings.

The scope of our model is limited to the "generalist" part of aesthetic judgment. More specifically, Immanuel Kant asserts two necessary properties of an aesthetic judgment:*(1) subjectivity* (being based on a feeling of pleasure and displeasure, rather than being empirical), and *(2) universality* (involving an expectation or claiming on the agreement of others) [40]. The contradicting nature of these properties lead to the "Big Question" of aesthetics: whether it is even possible for a subjective judgment to be universal [41]. The big question is the subject of an ongoing debate, where the *generalist* view holds that there exist general reasons for aesthetic judgments, and the *particularist* view denies that aesthetic judgments rely on general reasons.

The practice seems to be somewhere between both views: "each to their own taste" does apply to aesthetic judgments, but there is also a notable degree of agreement between judgments of different people in support of the generalist view. Subscribing to the generalist view enables a computational model of image aesthetics, but also draws the limits of such a model by ignoring artistic intention as well as previous knowledge and contextual information (Figure 2). Especially for the ultimate goal of correctly predicting a numeric overall aesthetic rating, this inherent limitation of automated image aesthetics poses an obstacle. Consequently, while we show that our method performs better than the state-of-the-art in predicting an overall aesthetics rating (Section 5.4), the focus of this work is on the design, computation and calibration of a meaningful aesthetic signature that summarizes representative photographic properties of an image.

Fig. 2. Artistic intention: focusing the camera on the subject is a common practice for shooting appealing photographs (eagle - left), but a photographer can intentionally break this common rule for artistic expression (bubbles - right). The generalist view of aesthetics captures only the former case. Images courtesy of Marina Cano (left) and Jeff Kubina (right)

3 Aesthetic Attributes

One of the main challenges of automated image aesthetics is identifying a set of aesthetic attributes that can be expressed algorithmically, and are closely related to photographic principles they claim to model. Since it is practically impossible that a computational system accounts for every photographic rule, one needs to determine some guidelines for choosing some aesthetic attributes over others. In this work, we considered the following criteria while determining the set of aesthetic attributes:

- **Generality:** while sharpness is relevant in every photograph, a more specific attribute such as facial expression is only useful for photographs with people. We chose not to limit our work to a specific type of photographs, and accordingly we selected among the more general attributes.
- **Relation to photographic rules:** from a modeling point of view it may be desirable that the aesthetic attributes are orthogonal to each other. However this would also require to invent new, artificial attributes that are not necessarily meaningful to humans, since in reality the photographic rules are not always orthogonal. In this work our main goal was to compute a multi-dimensional human interpretable aesthetic signature, and accordingly we chose to closely follow the photographic rules at the cost of possibly correlated attributes.

- **Clear definition:** in photography literature photographic rules and practices are often communicated through examples rather than mathematical formulas or concrete statements. For the purpose of automating image aesthetics we selected attributes that can be defined as clearly as possible.

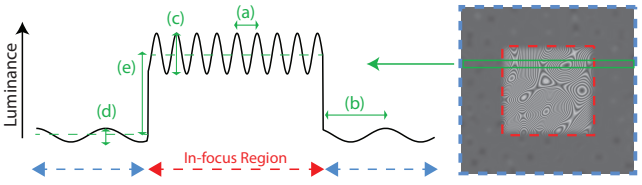

Fig. 3. From photographic rules to concretely defined aesthetic attributes: the 1D luminance (left) obtained by taking a slice from an abstract image (right) is used to build an intuition on how to express photographic rules in computational terms.

In the rest of this section we discuss the photographic rules selected based on the above principles, that form the foundation of the aesthetic attributes we use in our system. Using the 1D luminance profile in Figure 3-left obtained from the abstract image (right), we also investigate each rule in image processing terms to form a basis for our discussion later in Section 5.2. During our discussion of each photographic rule we highlight the relevant image features such as the spatial frequency of the in-focus region (a) and the background (b), the contrast magnitude of the in-focus region (c) and the background (d), and the luminance difference between the two regions (e) as depicted in Figure 3-left. The aesthetic attributes we discuss do not cover all aspects of image aesthetics, but are still expressive enough to enable multiple novel applications (Section 6). Moreover our framework can possibly be extended with other aesthetic attributes by following the workflow discussed in Sections 4 and 5.

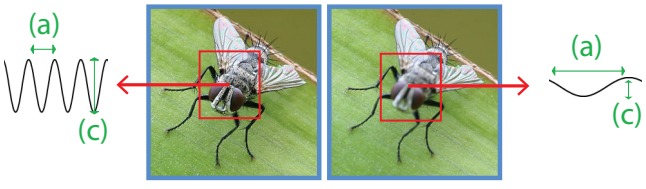

Fig. 4. A visual comparison between a sharp (left) and out-of-focus (right) photograph. Photographs courtesy of Muhammad Mahdi Karim.

In photography often times the camera is either focused to the entire scene, or to some specific scene object. An important rule of photography is ensuring that the in-focus region is **sharp** (Figure 4-left). Pictures with no scene elements in focus are often conceived as photographic errors. In fact, sharpening the in-focus region or the entire image is one of the very common post-processing operations to correct out-of-focus photographs, or to enhance the aesthetic quality of already sharp pictures. Sharpness is related to the magnitude and frequency of the image contrast within the in-focus region (Figures 4 and 3-a,c).

Fig. 5. A narrow depth-of-field significantly blurs background details (right) that would otherwise be in focus (left). Photographs courtesy of Henry Firus.

On the other hand, increasing the **depth** of the photograph through the use of specific camera lenses is a technique often employed by professional photographers (Figure 5-right). The presence of sharp, in-focus scene elements, together with other objects or background elements that are blurred enhances the depth impression of a photograph. In contrast, pictures with a wide depth-of-field and thus less variation in lens blur look flatter and often times less professional (Figure 5-left). Depth is related to the dominant contrast frequencies of the different image regions (Figures 5 and 3-a,b).

Fig. 6. The clarity rule favors photographs with a clearly identifiable region of interest and background (left), rather than visually cluttered photos where identifying a background and a center of interest is not even possible (right) Photographs courtesy of Miguel Lasa(left) and Julio Segura Carmona (right).

The **clarity** rule of photographic composition states that each picture should have a clear principal idea, topic, or center of interest to which the viewer's eyes are attracted [1]. Consequently, the use of empty regions (also called negative space) is considered as an important photographic skill (Figure 6-left). Shooting cluttered photographs with no obvious center of interest is one of the most common mistakes among amateur photographers (Figure 6-right). Photographic clarity is related both to the size of the in-focus region, as well as the contrast magnitudes of the in-focus region and the background regions (Figure 6 and 3-c,d).

The efficient use of the medium's dynamic range is another consideration when shooting photographs. As a general rule, if the difference in lightness between the brightest and darkest regions of a photograph is low, it can be perceived as under– or

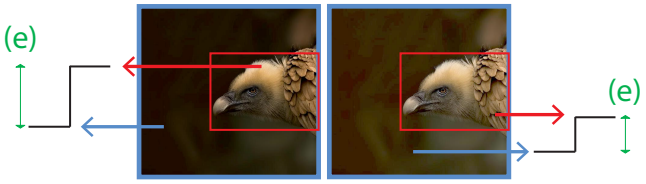

Fig. 7. Adjusting the camera settings to increase global contrast (left) often produces more plausible results than not utilizing the full dynamic range available (right). Photograph courtesy of Marina Cano.

over-saturated and "washed-out" (Figure 7-left). A sufficient level of global contrast often results in more plausible photographs (Figure 7-right). In fact, the **tone** attribute is related to the magnitude of the global lightness difference over the entire image (Figure 7 and 3-e).

In addition to these attributes, the chromatic information of the image also plays an important role in image aesthetics. The **colorfulness** attribute can be used to differentiate photographs with lively and saturated colors from photographs with desaturated colors.

During our discussion we indicate the usually preferred directions the photographic rules, but we refrain from making any general statements on how each rule affects the overall aesthetics rating of an image. Due to the "particularist" aspects of the aesthetic judgment process it is not uncommon that photographs not conforming to one or more of these rules are found aesthetically pleasing by the majority. Figure 6-right shows such an example that definitely violates the clarity rule, but in terms of overall aesthetics is still ranked among the highest in the photo.net dataset [3]. The next section discusses a subjective study where we obtained ground truth ratings for each aesthetic attribute. The ground truth data is later used to design and calibrate the aesthetic attribute metrics presented in Section 5.2.

4 Subjective Rating Study

In this section we present an experiment where we obtained subjective ratings for the aesthetic attributes on a set of real world images. Using this subjective data we designed and calibrated the aesthetic attribute metrics we describe in the next section.

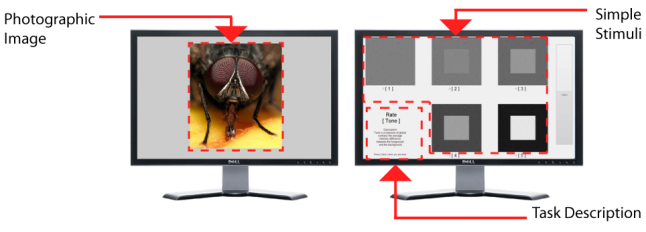

Fig. 8. An illustration of our experimental setup.

Experiment During our experiment the subjects were seated comfortably in front of two computer displays at a distance of approximately 0.5 meters. One of the displays showed a photographic image from our test set, whereas the other display showed 5 simple stimuli and a short task description (Figure 8). The task consisted of rating a *single* aesthetic attribute of the photographic image among sharpness, depth, clarity, tone and colorfulness on a 5-point scale. The simple stimuli generated separately for each aesthetic attribute were used to assist the subject by providing a neutral basis for each point of the rating scale.

Stimuli We assembled a test set that comprised 20 different images per aesthetic attribute, all obtained from the photo.net data set where each image had an overall aesthetic rating assigned by a community of semi-professional photographers (refer to Datta et al. [3] for an analysis of the images and ratings). The images used in our experiment were manually selected with an effort to maximize the diversity of the attribute ratings as well as the overall aesthetic ratings.

A common problem of subjective rating experiments in the absence of a reference is that the subject has often no baseline for assessing the measured effect. This often causes the earlier subjective responses to be unreliable until the subjects see the more extreme cases in the experimental test set, and use those as anchor points for their judgment during the remaining trials. While as a counter-measure such experiments are often preceded by a short training session, especially for highly subjective tasks as ours, it is highly desirable to additionally provide a baseline for rating without biasing the subject. This task is challenging, because one cannot simply use real-world photographs that would represent each rating on the 5-point scale, since the contents of the chosen photographs could invoke different reactions in different subjects and introduce unforeseen biases to the subjective ratings. To prevent this, we generated a set of 5 abstract images (one for each point in the rating scale) per aesthetic attribute building upon the abstraction in Figure 3. Our experiment was still preceded by a conventional training session, but additionally we used the emotionally neutral simple stimuli as a baseline for rating at each trial of the subjective experiment. Despite the presence of the simple stimuli, our subjects were made clear to ultimately rely on their own understanding of each attribute to prevent constraining their judgments.

The simple stimuli consisted of a square that region represents a foreground object, centered in a larger square that represents the background (Figure 9). A random texture pattern was generated separately for the foreground and the background using Perlin noise. The stimuli for sharpness and depth were generated by applying Gaussian blur to the foreground texture and background texture, respectively. The clar-

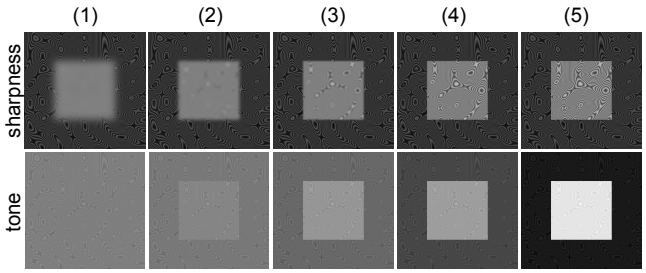

Fig. 9. Simple stimuli for sharpness and tone used in the subjective experiment to provide subjects a baseline for ranking. See supplemental material for the full set.

ity stimuli varied in the difference in the contrast magnitude of the texture between the foreground and the background, whereas the tone stimuli varied in the intensity difference between both image regions. On the other hand, the colorfulness stimuli was generated by modulating the saturation and size of a rainbow pattern. For all aesthetic attributes, the simple stimuli were generated to provide 5 roughly visually equal steps within the whole range of possible magnitudes. Each time when the subjects were evaluating an image in terms of an aesthetic attribute, the 5-level simple stimuli for the corresponding aesthetic attribute were presented on the side.

Procedure The photographs in our test set were presented in a random order. At each trial, the 21 subjects participated in our study were asked to rate a single aesthetic attribute of the presented image without any time constraints. All subjects had near-perfect or corrected eyesight and their ages ranged from 21 to 42. Among the subjects, 3 were asked to rate the entire set, whereas the others rated either the test images for sharpness, clarity and colorfulness, or the test images for depth and tone. Each subject participated in a training session that preceded the experiment where they were briefed on the experiment and the aesthetic attributes.

The results of our study is summarized in Figure 11-left, where we show the median ratings and subjective variation for each test image and aesthetic attribute. The following section discusses the aesthetic attribute metrics we designed and calibrated using the subjective ground truth data obtained in this section.

5 Metrics and their Calibration

In this section we present metrics for automatic computation of a subjectively meaningful rating for each aesthetic attribute we discussed previously. We first discuss the common components utilized by multiple metrics (Section 5.1), then present the mathematical formulation of each metric (Section 5.2), show the results of the calibration process where we map each metric's outcome to a perceptually meaningful 5-point scale (Section 5.3), and finally investigate the aesthetic judgment problem where we compare the accuracy of our metric with the state-of-the-art (Section 5.4). While we do not claim that the metrics presented in this section capture all the complexity of the subjective aesthetic analysis process, our final system has sufficient prediction power to enable multiple practical applications (Section 6).

5.1 Pre-processing Steps

In this section we discuss the computation of the core components of our system that are later utilized by the metrics we discuss in Section 5.2. Our metric takes an 8-bit RGB image as input, that is immediately converted to double precision and normalized. We start by computing an edge-stopping pyramid of the input image. The motivation for using an edge-stopping filter over a Gaussian was preventing contrast diffusion between the in-focus region and background, based on the assumption that the in focus region will likely be bordered by strong gradients. For each color channel I of the input image, each level $k \in [1, K]$ of the edge-stopping pyramid is defined as follows:

$$LP^k = dt(I, \sigma_r, \sigma_s, N), \tag{1}$$

where dt is the domain transform operator [42], the range parameter $\sigma_r = 1$ for $k < K$, $\sigma_r = Inf$ for $k = K$, and the spatial parameter $\sigma_s = 2^{(k+1)}$. In our implementation we set the number of pyramid levels to $K = 8$, and the number of domain transform iterations $N = 5$. One can use other edge preserving filtering methods instead of the domain transform, our choice was motivated by performance and is not crucial to the overall method.

We denote the absolute differences between the input image and LP^1 as the the detail layer D^1. The remaining detail layer levels D^k are defined as the differences between subsequent pyramid levels LP^k and LP^{k+1}. By combining all detail layers as follows, we compute a multi-scale contrast image:

$$C = \left|I - LP^1\right| + \sum_{k=1}^{K-1} \left|LP^k - LP^{k+1}\right|. \tag{2}$$

The detail layers D^k are also used to compute another data structure we call the "focus map". The focus map provides a rough spatial frequency based segmentation of the input image, that is later used by multiple aesthetic attribute metrics. Figure 10 shows an illustration of a focus map with 3 levels. The computation steps are as follows: we first apply the domain transform to each detail layer D^k using the gradients from the original image as the edge stopping criterion. We designate these filtered detail layers as $\hat{D}^k$ with $k = [1, K-1]$. The focus map levels F^k are computed sequentially through the following

formula:

$$F^k = \hat{D}^k \cdot \left[M \;\&\; \left(\hat{D}^k > \hat{D}^{k+1} \right) \right] \tag{3}$$

Where M is initialized as a matrix of ones with the same size as the image, and is updated at each level k as $M = M \;\&\; (\hat{D}^k == 0)$.

Some other computational elements required by the aesthetic attribute metrics are the following: we denote F^{if} as the binary map of the in-focus region and obtain it by marking each non-zero valued pixel of F^1 is as 1, and all zero valued pixels as 0. The inverse of F^{if}, denoted as F^{oof}, is the binary map of the remaining, out-of-focus regions. The area of the in-focus region A^{if} is defined as the number of all the non-zero pixels of the in-focus region F^{if} normalized by the number of total image pixels. The area of the out-of-focus region (A^{oof}) is computed similarly. In the following section we formulate the aesthetic attribute metrics by making use of these computational elements.

Fig. 10. Illustration of a 3-level "focus map". As an intermediate computational step we compute a spatial frequency based segmentation of the input image, that is later utilized by multiple aesthetic attribute metrics. Brighter colors of the focus map indicate stronger image contrast. Image courtesy of Bülent Erol.

5.2 Aesthetic Attribute Metrics

Our experience with different variations of aesthetic attribute metrics showed a tradeoff between metric sophistication and generality. While we were able to formulate complex metrics with high accuracy on specific small data sets, we also found that such metrics generate unpredictable results for previously unseen and/or larger data sets. Since generality is one of our main goals, the metrics we present in this section are rather simple and mostly direct implementations of the aesthetic attribute definitions from Section 3. As also discussed in Section 3, a consequence of directly implementing common rules of photography is that the predictions of our metrics are not necessarily linearly independent. However, by using such natural attributes, that in turn form the aesthetic signature, our method enables novel applications discussed in Section 6.

Equipped with the focus map (F), the multi-scale contrast image (C), the image luminance (L), and the area measures (A), the aesthetic attribute metrics are given in Table 1:

Sharpness ψ_{sh}	$\mu(\lvert F^1 \rvert)$
Depth ψ_{de}	$\underset{k}{\text{argmax}} \left[\sum (F^k > 0) \right] \; k = [2, K]$
Clarity ψ_{cl}	$A^{oof} \cdot (\lvert \mu(C \cdot F^{if}) - \mu(C \cdot F^{oof}) \rvert)$
Tone ψ_{to}	$c^u \cdot c^o \cdot \left\lvert p^{95}(L) - p^5(L) \right\rvert$
Colorfulness ψ_{co}	$f(I_r, I_g, I_b)$
$\mu(A)$	$\frac{1}{N} \sum_{x=1}^{N} A(x), \quad A(x) > 0$
p^n	n-th percentile

TABLE 1
Equations of the aesthetic attribute metrics. Refer to text for details.

The **sharpness** metric consists of the average absolute contrast magnitude at the first level of the focus map (F^1). The outcome of this metric increases with stronger high frequency details over a larger area. We found that the average contrast measure is more stable than taking the maximum, mostly due to the effect of very small areas with exceptionally sharp details.

The **depth** metric estimates the perceived range of blur among all image details by searching the remaining focus map levels for the largest area with a non-zero response. The resulting index of the focus map level indicates the dominant spatial frequency at the non-sharp image regions. The metric predits zero depth for completely out-of-focus images. As an example, Figure 10 shows a 3-level focus map, where the first level contains sharp details in focus, and the remaining two levels contain the out-of-focus regions. In this example the metric would choose the second level of the focus map, since most of the image's non-sharp regions are at that level.

The **clarity** metric utilizes two terms: the area of the out-of-focus region, and the difference of average contrast within the in-focus region with the average contrast in the out-of-focus regions. The metric outcome increases with the presence of large image regions that are either empty or contain low contrast details.

The **tone** metric computes the difference of the maximum and minimum gamma corrected luminance values of the image. Perceptual studies on lightness perception have shown that utilizing the extreme values is preferable over using the mean luminance [43]. However, to eliminate the error from isolated pixels with very high and low luminance values, we compute the 95^{th} percentile and 5^{th} percentile instead of maximum and minimum. To account for under– and over-exposure we introduce two additional terms c^u and c^o. These terms are defined as $c^u =$

$min(u, p^{30}(L) - p^{5}(L))/u$ and $c^o = min(o, p^{95}(L) - p^{70}(L))/o$. The terms penalize large image regions with significantly low or high luminance, and therefore are likely to be under– or over-exposed. We set u and o to the pixel values 0.05 in our implementation.

The **colorfulness** metric utilized in our framework is discussed in detail in Hasler et al [44]. Given an RGB image, this method estimates colorfulness by standard deviation and mean of the opponent color channels yellow-blue and red-green. Our choice of this metric was motivated by the fact that it has been perceptually validated on a set of complex images. Nevertheless, we re-evaluate the method in our calibration experiment discussed in the next section.

5.3 Calibration

The metrics we discussed in Section 5.2 do not necessarily have the same range of output values. Presenting the aesthetic attribute ratings together in the aesthetic signature format requires that they are calibrated with subjective responses. This ensures both that each individual attribute rating matches the corresponding subjective rating, and also ratings of different attributes are normalized to the same range. To that end we calibrate the aesthetic attribute metrics using the subjectively obtained aesthetic attribute ratings from Section 4 (Figure 11-left). The calibration procedure involves for each aesthetic attribute building a linear system of equations from the metric outcomes and the median values of subjectively obtained ground truth ratings. By solving the equation system we determine the multipliers for calibration. Figure 11-right shows the subjective ground truth ratings with the calibrated metric outcomes computed automatically by our system.

The figure shows that given the highly subjective nature of the aesthetic attributes our metrics provide a good estimate to the ground truth data. The root mean square error of the calibrated metric predictions with respect to the subjective data are 0.34 for sharpness, 0.32 for depth, 0.31 for tone, 0.33 for clarity and 0.26 for colorfulness. Note that the relatively lower colorfulness error can be misleading, since 4 images in our test set were grayscale, and thus predicting the colorfulness rating was trivial. Excluding these images, the accuracy of the colorfulness metric is at the same level as the other aesthetic attribute metrics. As mentioned in Section 5.2 one can design more complex metrics to improve on prediction accuracy at the cost of generality. However we found that this approach limits the practical utility of the system, and the current performance is sufficient to enable a number of novel applications without sacrificing generality (Section 6). Note that our metrics are trained with natural 8-bit images, obtaining reliable aesthetic predictions for synthetic and/or HDR images would likely require additional training and modifications to the metrics.

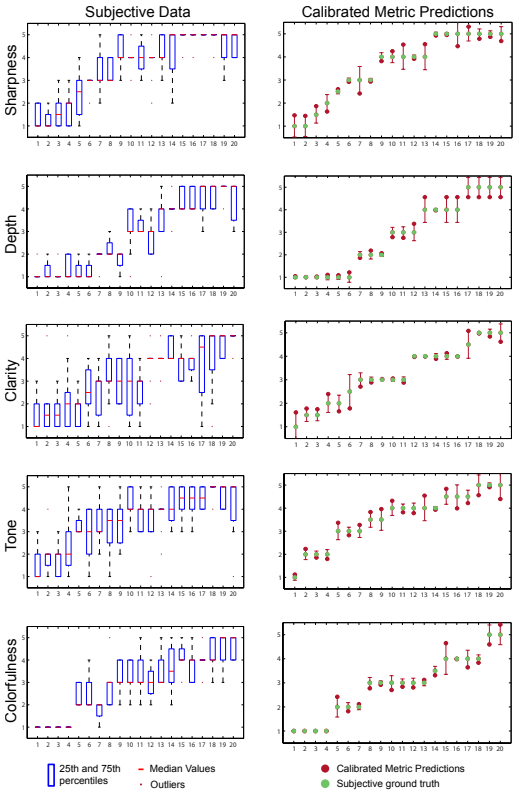

Fig. 11. Experimentally obtained subjective aesthetic attribute ratings (left column), and the predictions of our calibrated metrics where error bars show the accuracy with respect to the subjective ground truth (right column).

5.4 Aesthetic Judgment

Having computed the calibrated ratings for each individual aesthetic attribute, an obvious question is if we can combine them to obtain an overall aesthetics rating of an image. As we discussed earlier, this is a challenging problem due to the "particularist" aspects of aesthetic judgment (Section 2). While aesthetic judgment in this form is not the main focus of this work, we conducted a study both to investigate how our method generalizes beyond the calibration set, and to compare our method's prediction accuracy with *acquine*, a popular online aesthetic rating engine [16]. We automatically collected a set of 955 images from the dpchallenge dataset [3], which contains photographic images with associated subjective overall aesthetic ratings, and has often been used as ground truth data to train and evaluate automated image aesthetics methods.

For generating the test set, the obvious approach of randomly sampling from the entire dpchallenge set would have produced a bias towards images with medium overall aesthetic ratings, since the number of very high and very low ranked images in the set are much less than the number of images with medium ratings [3]. We produced a balanced test set

by dividing the range of overall aesthetics ratings into equally sized intervals and sampling the images closest to the center of each interval.

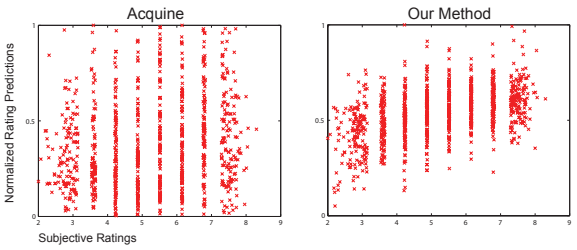

Fig. 12. A comparison of the overall aesthetic rating predictions with the subjective ground truth data. Our method's (right) correlation to the ground truth is significantly higher than that of acquine's (left).

Figure 12-left shows each image's acquine rating plotted against the subjective overall rating from the ground truth data. The correlation between the acquire rating and the ground truth is 0.17. On the same set we also computed each image's aesthetic signature, which we aggregated into an overall aesthetics rating using the following simple formula:

$$\omega = \psi_{sh} \times \mu\left(\psi_{de}, \psi_{cl}, \psi_{to}, \psi_{co}\right), \qquad (4)$$

where μ denotes the mean function. Equation 4 is motivated by our observation that sharpness is *essential* to an aesthetically pleasing photograph, in that its absence will result in a low overall aesthetics rating. This is not necessarily the case for the other attributes: e.g. a grayscale photograph can still be considered as aesthetically pleasing, if it has high ratings in clarity, depth and tone. As shown in Figure 12-right our method's predictions are noticeably more accurate than acquine's predictions. In fact, the correlation between ω (Equation 4) and the ground truth is considerably higher at 0.46. Thus, even though Equation 4 is admittedly ad-hoc, it is justified by the fact that it achieves significantly better correlation to the subjective data compared to the state-of-the-art.

Despite our method's better accuracy, the results of this study show that there is still a lot room for improvement for the task of predicting an overall aesthetics rating. The correlation between our overall aesthetic rating prediction and the subjective data suggests that there is indeed a link between the individual aesthetic attributes and the overall aesthetic judgement. However it is important to note that due to the highly subjective nature of the aesthetic judgement process, more validation would be necessary in order to be able to claim that any automatic aesthetic judgement method works universally.

On the other hand the aesthetic signature representation not only gives an impression on overall aesthetics, but also provides some insight on what makes an image aesthetically plausible or not as demonstrated

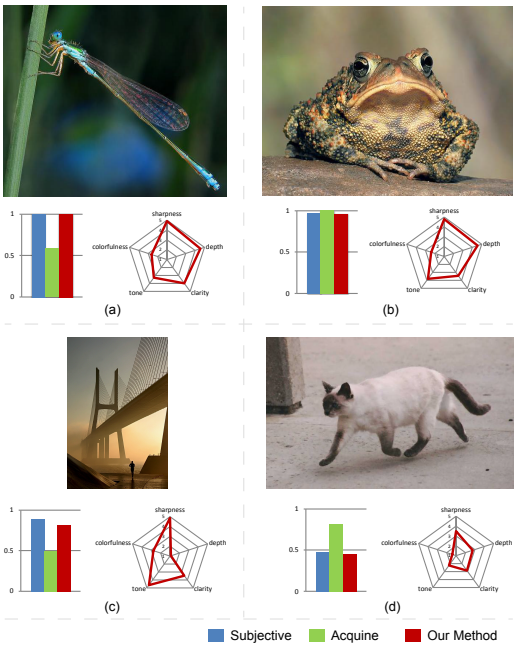

Fig. 13. Our method not only is more accurate in predicting an overall aesthetics score, but also provides some insight on why an image is aesthetically pleasing or not. See text for discussion.

in Figure 13. The figure compares the overall subjective aesthetic ratings, the acquine predictions, and the predictions obtained by applying Equation 4 along with the aesthetic signatures for a set of 4 exemplary images.

For these images, in addition to our overall aesthetics prediction being more accurate, our method enables a more detailed aesthetic evaluation through aesthetics signatures. For example in Figure 13, even though images (a) and (b) have completely different subjects, from a photography point of view they are similar since they both are macro photographs with a sharp foreground and a shallow depth-of-field. Image (a) makes better use of negative space and ranks slightly higher in clarity, whereas its global contrast is slightly lower reflected in its tone ranking. Image (c) on the other hand has a wide depth-of-field and very low lens blur resulting in low depth rating, but has stronger global contrast compared to (a) and (b) paralelled by its high tone rating. Image (c) is aesthetically pleasing just as images (a) and (b) are, but in a different way as indicated by the different shape of its aesthetic signature. Image (d), however, doesn't have anything that makes it special which is also evident from its aesthetic signature.

6 Applications

In this section we demonstrate various photo editing applications of the aesthetic signatures, such as automated aesthetic analysis, HDR tone mapping eval-

uation, and providing aesthetic feedback for multi-scale contrast manipulation. All results in this section are generated by our prototype system using standard model parameters discussed previously. Our current matlab based implementation requires roughly a second to compute the calibrated aesthetic signature of a 1 megapixel image on a desktop computer.

6.1 Automated Aesthetic Analysis

Figure 14 shows a set of pictures that are edited in various ways to make them visually more appealing. We compute the aesthetic signatures of the original and edited images separately. An advantage of our method is that it does not require pixel correspondence between the original and edited images, and also works if the editing involves cropping the original. A comparison of the resulting aesthetic signatures provides information on the nature and magnitude of the aesthetic improvements made in each case. For example, the original image (a) is dramatically improved in sharpness, depth, and to a lesser degree colorfulness. Image (c) has been also been improved in sharpness and colorfulness at the cost of slight decrease in clarity. The edited image has high contrast details over the entire image, and as a result no clearly identifiable subject region. The edited image (d) on the other hand is simply converted to grayscale with a slight increase in sharpness.

6.2 HDR Tone Mapping Evaluation

As a consequence of the numerous HDR tone mapping techniques in the literature, evaluation of the operators has received some interest in the research community. Subjective evaluation studies [45] are often prohibitively expensive due to the number of dimensions involved in such studies: test images, tone mapping operators, parameters, to name a few. Automated studies have been made possible through metrics such as DRI-IQM [46] that computes the probability of contrast loss, amplification and reversal. DRI-IQM evaluates tone mapping in terms of the amount of visible details preserved in the tone mapped image with respect to the reference HDR image. Figure 15 shows that our method can complement current tone mapping evaluation methods from an aesthetics perspective. Instead of comparing the tone mapped images with reference HDR images, our method directly evaluates tone mapped images produced by the operators Drago [47], Tumblin [48] and Ward [49]. The figure shows that for the tested set of images, produced by the authors of each operator, Ward's operator consistently results in more tone and sharpness compared to the other two. Another interesting result is the consistency of the aesthetic consequences of the tone mapping operators over multiple images. Figure 16 illustrates that the ratio of aesthetic signatures of two operators computed over a number of images can be used to characterize each tone mapping operator.

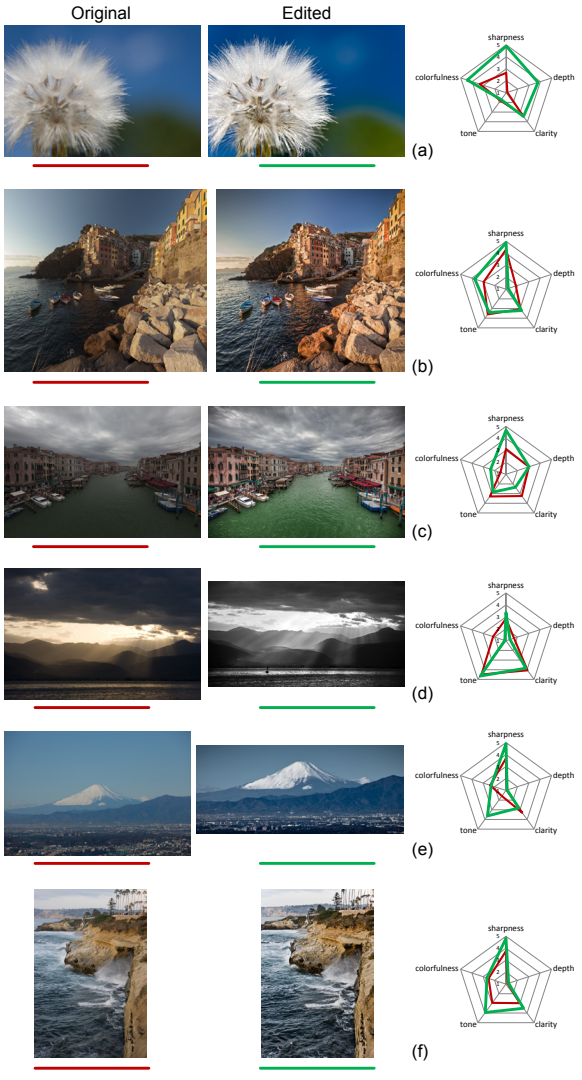

Fig. 14. The aesthetic changes before and after editing reflected in the automatically generated aesthetic signatures. Photographs courtesy of Wojciech Jarosz.

6.3 Multi-scale Contrast Editing

Recent work on multi-scale image editing has been focused on using efficient edge-preserving image decompositions on multiple scales, such as Farbman et al. [50], Fattal [51], Paris et al. [52], Gastal et al. [42] among many others. These frameworks provide better control over an image's aesthetic attributes compared to global operations like brightness change (Figure 17 - top row). Multi-scale frameworks allow enhancing image details at different frequencies by local contrast editing, through which one can enhance tone and sharpness without significant negative consequences if not taken to extremes (Figure 17 - rows 2 and 3). A possible use scenario of the aesthetic signatures is providing aesthetic feedback to the user during a

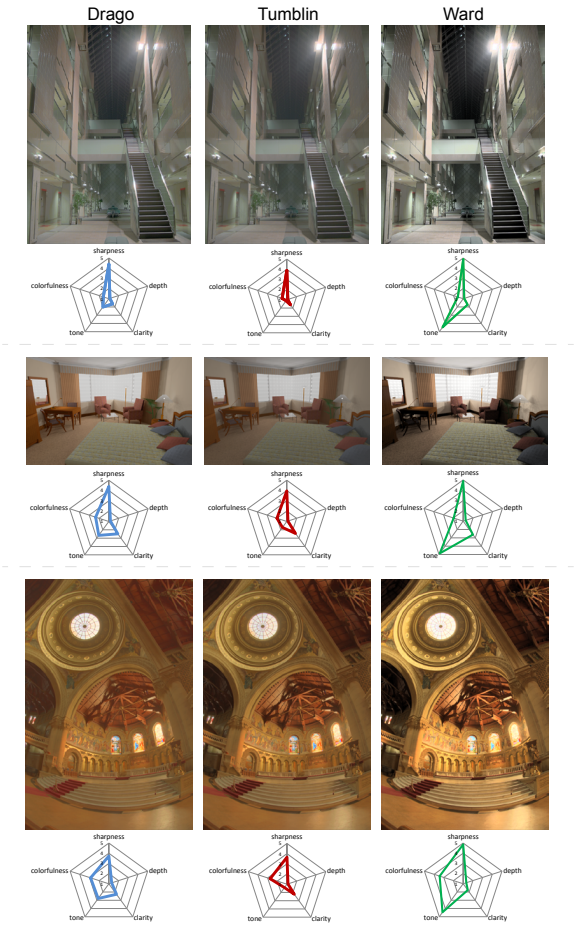

Fig. 15. Analysis of tone mapped HDR images using Drago's, Tumbin's and Ward's operators provided by their authors. Using our method one can analyze aesthetic attributes of the tone mapped images without requiring the reference HDR image.

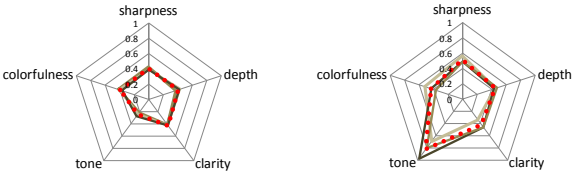

Fig. 16. The relative aesthetic signature of Drago's operator (left) and Ward's operator (right), obtained by normalizing the aesthetic signatures of both by the aesthetic signature of Tumblin's operator. The brown shades show the relative aesthetic signatures for the three images in Figure 15, and the red dotted lines show their average. Note that the relative aesthetic signatures of both operators are very similar over multiple images. This suggests that our method captures some essential properties of these operators invariant over multiple images.

multi-scale image editing session. This way, the user can be made aware of the aesthetic consequences

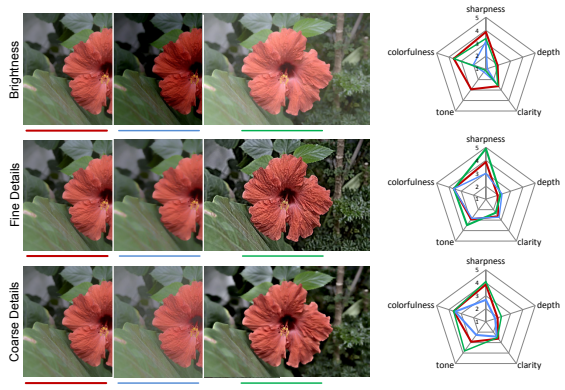

Fig. 17. Aesthetic consequences of some basic contrast editing operations. Image courtesy of Zeev Farbman.

of the changes he introduces to an image. In such a setting, our system can also prevent users taking certain actions that will make the image worse. Figure 18 shows two examples where both images have initially a high sharpness rating. As such, enhancing fine image details does not provide any aesthetic benefit in the first case (top row). In the second case (bottom row) the fine detail enhancement significantly reduces the aesthetic appeal of the image by completely negating the depth-of field effect, reflected by the reduced depth and clarity ratings.

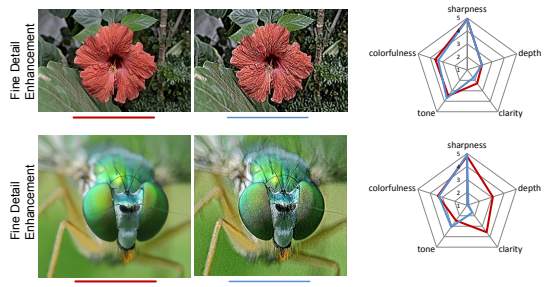

Fig. 18. Unnecessary editing can have negative aesthetic consequences. Since both top and bottom image have a high sharpness rating to begin with, enhancing fine image details does not provide any benefit (top row) or makes the image considerably less appealing (bottom row). Images courtesy of Zeev Farbman Mark Plonsky.

7 Conclusion

In this work we present a framework for automated image aesthetics that can be utilized in current photo editing software packages for providing aesthetic guidance to the user. Given a single image, our system automatically computes an aesthetic signature that comprises calibrated ratings of a set of aesthetic attributes, and provides a compact representation of

some of the input image's fundamental photographic properties. We described an experimental procedure for obtaining subjective ratings for each aesthetic attribute, and presented a set of metrics capable of accurately predicting these subjective ratings. We showed that our method outperforms the current state-of-the-art in predicting an overall aesthetics rating, as well as opening novel application areas.

Our aim was to be able to evaluate all types of real-world pictures without putting any constraints on their content or aesthetic characteristics, and the aesthetic attributes we utilized in this work were chosen accordingly. Consequently our method is limited to the expressiveness of the five aesthetic attributes we evaluate. While we showed our current system enables novel applications (Section 6), the aesthetic signature representation can still be enhanced by increasing its number of dimensions. Also, the aesthetic attribute metrics in our system trade off complexity for generality, because we did not want our metrics to overfit our subjective dataset, but rather to capture consistently present tendencies. This design choice also resulted in a slightly higher deviation of the metric predictions from subjective ground truth data than what we would have achieved with more complex metrics.

Since automated image aesthetics is fundamental to many visual computing methods, future directions for our work include in-depth treatment of novel application areas enabled by our method, such as the ones we demonstrated in Section 6. Another direction could be specializing our method by taking into account further, more specific aesthetic attributes like facial expressions, the position of the horizon line, rule of thirds, etc., that are only meaningful for certain types of photographs. Since in this work our reference data consists solely of real world images, another immediate future direction is testing to what extent our framework generalizes to synthetic images. It is also an interesting research question if there is a connection between recent work on image memorability [53] and aesthetics.

ACKNOWLEDGMENTS

Acknowledgements and image credits were removed for anonymity during the review cycle.

REFERENCES

[1] D. Freelan, *Photography (Basic), NAVEDTRA 14209*. Naval Education and Training Professional Development and Technology Center, 1993.

[2] A. E. Savakis, S. P. Etz, and A. C. P. Loui, "Evaluation of image appeal in consumer photography," in *Proc. of SPIE: Human Vision and Electronic Imaging*, vol. 3959, no. 1, 2000, pp. 111–120.

[3] R. Datta, J. Li, and J. Z. Wang, "Algorithmic inferencing of aesthetics and emotion in natural images: An exposition," in *Proc. of IEEE Int. Conf. on Image Processing*, 2008, pp. 105–108.

[4] D. Joshi, R. Datta, E. Fedorovskaya, Q.-T. Luong, J. Wang, J. Li, and J. Luo, "Aesthetics and emotions in images," *Signal Processing Mag., IEEE*, vol. 28, no. 5, pp. 94 –115, 2011.

[5] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in *Proc. of ECCV*, 2006, pp. 7–13.

[6] A. K. Moorthy, P. Obrador, and N. Oliver, "Towards computational models of the visual aesthetic appeal of consumer videos," in *Proc. of European Conf. on Computer vision (ECCV)*, 2010, pp. 1–14.

[7] Y. Ke, X. Tang, and F. Jing, "The design of high-level features for photo quality assessment." in *In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2006, pp. 419–426.

[8] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka, "Assessing the aesthetic quality of photographs using generic image descriptors," in *IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 1784 –1791.

[9] Y. Luo and X. Tang, "Photo and video quality evaluation: Focusing on the subject," in *Proceedings of the 10th European Conference on Computer Vision (ECCV)*, 2008, pp. 386–399.

[10] C. Li and T. Chen, "Aesthetic visual quality assessment of paintings," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 3, no. 2, pp. 236–252, 2009.

[11] C. Li, A. C. Loui, and T. Chen, "Towards aesthetics: a photo quality assessment and photo selection system," in *Proceedings of the international conference on Multimedia*, 2010.

[12] L. Yao, P. Suryanarayan, M. Qiao, J. Wang, and J. Li, "Oscar: On-site composition and aesthetics feedback through exemplars for photographers," *International Journal of Computer Vision*, pp. 1–31, 2011.

[13] C.-H. Yeh, Y.-C. Ho, B. A. Barsky, and M. Ouhyoung, "Personalized photograph ranking and selection system," in *ACM Multimedia 2010*, October 2010, pp. 211–220.

[14] X. Sun, H. Yao, R. Ji, and S. Liu, "Photo assessment based on computational visual attention model," in *Proceedings of ACM international Conf. on Multimedia*, 2009, pp. 541–544.

[15] P. Obrador and N. Moroney, "Evaluation of image appeal in consumer photography," in *Proc. of SPIE: Image Quality and System Performance*, vol. 7242, no. 1, 2009, pp. (72 420T) 1–12.

[16] R. Datta and J. Z. Wang, "Acquine: aesthetic quality inference engine - real-time automatic rating of photo aesthetics." in *Multimedia Information Retrieval*, 2010, pp. 421–424.

[17] S. Bhattacahrya, R. Sukthankar, and M. Shah, "A framework for photo-quality assessment and enhancement based on visual aesthetics," in *Proc. of the international conference on Multimedia*, 2010, pp. 271–280.

[18] Y.-J. Liu, X. Luo, Y.-M. Xuan, W.-F. Chen, and X.-L. Fu, "Image retargeting quality assessment," *Computer Graphics Forum (Proc. of Eurographics)*, vol. 30, no. 2, pp. 583–592, 2011.

[19] L. Liu, Y. Jin, and Q. Wu, "Realtime aesthetic image retargeting," pp. 1–8, 2010.

[20] L. Liu, R. Chen, L. Wolf, and D. Cohen-Or, "Optimizing photo composition," *Computer Graphics Forum*, vol. 29, 2010.

[21] H.-H. Su, T.-W. Chen, C.-C. Kao, W. Hsu, and S.-Y. Chien, "Preference-aware view recommendation system for scenic photos based on bag-of-aesthetics-preserving features," *Multimedia, IEEE Transactions on*, vol. 14, no. 3, pp. 833–843, 2012.

[22] P. O'Donovan, A. Agarwala, and A. Hertzmann, "Color compatibility from large datasets," *ACM Transactions on Graphics (Proc. of SIGGRAPH)*, vol. 30, no. 4, 2011.

[23] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y.-Q. Xu, "Acm transactions on graphics (proc. of siggraph)," *Color harmonization*, pp. 624–630, 2006.

[24] M. Nishiyama, T. Okabe, I. Sato, and Y. Sato, "Aesthetic quality classification of photographs based on color harmony," in *Proc. of CVPR*, 2011, pp. 33–40.

[25] J. Fiss, A. Agarwala, and B. Curless, "Candid Portrait Selection From Video," *ACM Trans. on Graphics*, vol. 30, no. 6, 2011.

[26] S. Dhar, V. Ordonez, and T. L. Berg, "High level describable attributes for predicting aesthetics and interestingness," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 1657–1664.

[27] F. Perronnin, "Ava: A large-scale database for aesthetic visual analysis," in *IEEE CVPR*, 2012, pp. 2408–2415.

[28] S. Daly, "The Visible Differences Predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*. MIT Press, 1993, pp. 179–206.
[29] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "Hdr-vdp-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions," vol. 30, pp. 40:1–40:14, 2011.
[30] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. Morgan & Claypool Publishers, 2006.
[31] C. Chen, W. Chen, and J. A. Bloom, "A universal reference-free blurriness measure," in *SPIE vol. 7867*, 2011.
[32] S. Daly and X. Feng, "Decontouring: Prevention and removal of false contour artifacts," in *Proc. of Human Vision and Electronic Imaging IX*, ser. SPIE, vol. 5292, 2004, pp. 130–149.
[33] S. Winkler, "Perceptual video quality metrics – a review," in *Digital Video Image Quality and Perceptual Coding*, H. . R. . Wu and K. . R. . Rao, Eds. CRC Press 2006, 2006, pp. 155–179.
[34] H. Sheikh, A. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. on Image Processing*, vol. 14, no. 11, pp. 1918 –1927, 2005.
[35] J. Kopf, W. Kienzle, S. Drucker, and S. B. Kang, "Quality prediction for image completion," *ACM Trans. on Graphics (Proc. of SIGGRAPH)*, vol. 31, no. 6, pp. 131:1–131:8, 2012.
[36] P. Whittle, "Increments and decrements: Luminance discrimination," *Vision Research*, vol. 26, no. 10, pp. 1677–1691, 1986.
[37] G. Ramanarayanan, J. Ferwerda, B. Walter, and K. Bala, "Visual equivalence: towards a new standard for image fidelity," *ACM Transactions on Graphics (Proc. of SIGGRAPH)*, vol. 26, 2007.
[38] G. Ramanarayanan, K. Bala, and J. A. Ferwerda, "Perception of complex aggregates," *ACM Transactions on Graphics (Proc. of SIGGRAPH)*, vol. 27, pp. 60:1–60:10, 2008.
[39] J. Křivánek, J. A. Ferwerda, and K. Bala, "Effects of global illumination approximations on material appearance," *ACM Trans. on Graphics (Proc. of SIGGRAPH)*, vol. 29, pp. 112:1–10, 2010.
[40] D. Burnham. (2001) Internet encyclopedia of philosophy - Kant's aesthetics. [Online]. Available: http://www.iep.utm.edu/kantaest
[41] N. Zangwill, "Aesthetic judgment," in *The Stanford Encyclopedia of Philosophy*, 2010.
[42] E. S. L. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Transactions on Graphics (Proc. of SIGGRAPH)*, vol. 30, no. 4, pp. 69:1–69:12, 2011.
[43] G. Krawczyk, K. Myszkowski, and H.-P. Seidel, "Lightness perception in tone reproduction for high dynamic range images," in *Proc. of Eurographics*, vol. 24, no. 3, 2005.
[44] D. Hasler and S. Süsstrunk, "Measuring colourfulness in natural images," in *Proc. of SPIE: Human Vision and Electronic Imaging*, vol. 5007, 2003, pp. 87–95.
[45] A. Yoshida, V. Blanz, M. K., and H.-P. Seidel, "Perceptual evaluation of tone mapping operators with real-world scenes," in *Proc. of SPIE: Human Vision and Electronic Imaging*, 2005, pp. 192–203.
[46] T. O. Aydın, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Dynamic range independent image quality assessment," in *ACM Trans. on Graphics (Proc. of SIGGRAPH)*, vol. 27(3), 2008.
[47] F. Drago, K. Myszkowski, T. Annen, and N.Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Computer Graphics Forum (Proc. of EUROGRAPHICS)*, vol. 22, no. 3, 2003.
[48] J. Tumblin, J. K. Hodgins, and B. K. Guenter, "Two methods for display of high contrast images," *ACM Transactions on Graphics*, vol. 18, no. 1, pp. 56–94, January 1999, iSSN 0730-0301.
[49] G. Ward Larson, H. Rushmeier, and C. Piatko, "A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes," *IEEE Transactions on Visualization and Computer Graphics*, vol. 3, no. 4, pp. 291–306, 1997.
[50] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," in *ACM Transactions on Graphics (Proc. of SIGGRAPH)*, vol. 27, no. 3, 2008.
[51] R. Fattal, "Edge-avoiding wavelets and their applications," *ACM Transactions on Graphics (Proc. of SIGGRAPH)*, vol. 28, no. 3, 2009.
[52] S. Paris, S. W. Hasinoff, and J. Kautz, "Local laplacian filters: edge-aware image processing with a laplacian pyramid," *ACM Trans. on Graphics (Proc. of SIGGRAPH)*, vol. 30, no. 4, 2011.
[53] A. Khosla, J. Xiao, A. Torralba, and A. Oliva, "Memorability of image regions," in *Advances in Neural Information Processing Systems (NIPS)*, Lake Tahoe, USA, December 2012.

Tunç Ozan Aydın is an Associate Research Scientist at Disney Research Zürich. His main research interest lie in modelling various aspects of the human visual system, and applying these models to computer graphics and vision problems. He holds a PhD (summa cum laude) degree from the Computer Graphics Department of Max-Planck-Institut für Informatik (2010). His dissertation received the Eurographics PhD Award (2012). He obtained his M.S. degree from the College of Computing of Georgia Institute of Technology (2005), and his B.S degree from the Civil Engineering Department of Istanbul Technical University (2003).

Aljoscha Smolic is a Senior Research Scientist at Disney Research Zürich leading the Advanced Video Technology Group. Aljoscha Smolic received his Dipl.-Ing. degree in electrical engineering from the Technical University of Berlin, Germany, in 1996, and his Dr.-Ing. degree in electrical and information engineering from Aachen University of Technology (RWTH), Germany, in 2001. Dr. Smolic received the Rudolf-Urtlel-Award of the German Society for Technology in TV and Cinema (FKTG) for his dissertation in 2002. He is the Area Editor for Signal Processing: Image Communication and served as Guest Editor for the Proceedings of the IEEE, IEEE Transactions on CSVT, IEEE Signal Processing magazine, and other scientific journals.

Markus Gross is a Professor of Computer Science at the Swiss Federal Institute of Technology Zrich (ETH), head of the Computer Graphics Laboratory, and the Director of Disney Research, Zrich. He joint the ETH Computer Science faculty in 1994. His research interests include physically based modeling, computer animation, immersive displays, and video technology. Before joining Disney, Gross was director of the Institute of Computational Sciences at ETH. He received a master of science in electrical and computer engineering and a PhD in computer graphics and image analysis, both from Saarland University in Germany in 1986 and 1989. Gross serves on the boards of numerous international research institutes, societies, and governmental organizations. He received the Technical Achievement Award from EUROGRAPHICS in 2010 and the Swiss ICT Champions Award in 2011. He is a fellow of the ACM and of the EUROGRAPHICS Association and a member of the German Academy of Sciences Leopoldina as well as the Berlin-Brandenburg Academy of Sciences and Humanities. He receives a Technical Achievement Award from the Academy of Motion Picture Arts and Sciences in 2013. Prior to his involvement in Disney Research he cofounded Cyfex AG, Novodex AG, LiberoVision AG, and Dybuster AG.