Computational Aesthetics in Graphics, Visualization, and Imaging (2012), pp. 1–8
D. Cunningham and D. House (Editors)

# Evaluating Visual Aesthetics in Photographic Portraiture

Shehroz S. Khan and Daniel Vogel

David R. Cheriton School of Computer Science
University of Waterloo
{s255khan,dvogel}@uwaterloo.ca

**Abstract**

*We propose and demonstrate a strategy to quantify aesthetic quality in photographs. Our approach is to develop a small set of classification features by tuning general compositional principles to a targeted image domain where saliency can be better understood. We demonstrate this strategy with photographic portraits of individuals, but it can be extended to other domains. Our technique leverages a refined method of using templates as spatial composition feature look-up tables. Compared to the traditional approach using a large set of global and local features extracted with little salient knowledge, classifiers using features extracted with our approach are better predictors of human aesthetic judgments.*

## 1. Introduction

The combination of smart phones and social networking websites make it easy for anyone to take photographs and make them available to a wide audience. The problem is that casual photographers are not always good at assessing the aesthetic quality of their photographs. Harnessing the computational capability of a camera-equipped smart phone, an algorithm could offer an instant critique of the photograph based on aesthetic criteria. This may encourage the casual photographer to re-take the photo to improve its aesthetics, or reconsider whether it is a good candidate for sharing. However, algorithmic assessment of the aesthetic quality of a photograph remains a vexing problem.

### 1.1. Motivation

A challenge for effective aesthetic assessment is the difficult-to-model subjective aspect influenced by past experience, taste, and cultural context. However, there exist fundamental composition principles that improve the quality of a photograph. The challenge is to effectively apply these principles, one really needs to know what is in the image, i.e. the salient regions. Only with an understanding of the size and shape of these regions, the spatial relationships between regions, and the identification of what each region represents, can aesthetic quality be effectively assessed, even at level of adherence to compositional principles. Moreover, compositional principles are not truly universal [CL09]. For example, the often used rule-of-thirds [Chi08, Ric05] may apply differently to a landscape and a portrait — locating the face on one of the lower power-points is not desirable, but locating a tree in that position would be fine. Knowledge of what type of photograph is being assessed can be treated as a global saliency problem. Although there are promising results [LWT11], in general, algorithmic solutions for detecting saliency remain difficult.

Regardless of having little knowledge of image saliency, many researchers seek to quantify aesthetic quality by treating a photograph, or more generally an image, as an independent whole [DJLW06, BSS10, NOSS11]. The idea is to calculate many global features using a statistical analysis of the image, and approximate salient regions with basic segmentation algorithms to calculate local statistical features. The strategy behind this traditional approach is that a machine learning classifier will determine which of the many features are actually relevant when making an aesthetic assessment. This approach is essentially *searching* for a way to measure aesthetic quality, rather than *understanding* principles which improve aesthetic quality.

### 1.2. Contribution

Most of the previous studies have evaluated visual aesthetics across all images which could be landscapes, cityscapes, group portraits, single portraits, animals, product shots, or anything else [KTJ06, DOB11, LWT11]. This makes the application of compositional principles problematic. We believe recent work examining aesthetics in constrained im-

age domains, such as photographs of people [LGLC10], are a step in the right direction. But, even this is not adequately constrained. We examine a very targeted image space of photographic portraits of individuals. Not only does this make saliency detection more tractable, but targeted compositional principles may be applied specific to the targeted image domain. Moreover, the proliferation of smart phones and social media websites make photographs of individuals a relevant space.

Our approach leverages a top-down understanding of composition principles in portrait photographs. We identify a small set of 7 features which utilize the knowledge of salient face and background regions. We utilize a refined template-based feature extraction method based on Obrador et al. [OSHO10] for spatial composition and a small set of colour-space features for highlight and shadow composition. Compared to a traditional approach using 66 global and local statistical features, our approach improves accuracy by 4% to 6 % yet uses $1/9^{th}$ as many features.

The rest of the paper is organized as follows. Section 2 presents a background of visual aesthetics including basic composition principles of photographic portraiture and a review of past work assessing aesthetics in photographs. In Section 3, we describe two sets of features for assessing portrait photographs: a large set of global and local features commonly used in the bottom-up approach of past work, and our set of targeted features used in our top-down approach. Section 4 describes the design of our experiments and the results obtained. Section 5 concludes with directions for future work.

## 2. Background and Literature Review

### 2.1. Composition Principles in Photography

Artistic photography may follow no rules in the general sense, but following basic composition principles can make casual photographs more visually compelling and pleasing to look at. Spatial composition is a particularly important aspect as it relates to Gestalt theory. Krages [Kra05] explains this relationship as *the mind perceives the whole image without having to first analyze the parts*.

A common spatial composition principle is the rule-of-thirds, where the subject is placed along, or at the intersection of "power lines" that divide the image into thirds horizontally and vertically. With portraiture, the subject is the face and the rule-of-thirds can be refined to prioritize the top power line to keep the eyes in the upper third of the image area [Dic08] (Figure 1a). Simply centering an object in the frame achieves a symmetrical composition, and with portraiture this is arguably appropriate. The rule-of-thirds power lines can provide additional vertically shifted centers (Figure 1b).

Dickson [Dic08] and others [Chi08, Ric05] provide other tips for good composition in portrait photography that include using light and shadow to define the face (Figure 1c) and creating contrast between the person and the background (Figure 1d). Of course, there are many more of these sort of rules, but the aim of our work is to show that even using a small number of rules to create targeted features is more effective than trying to find classification patterns in a large set of predominantly global statistical features.

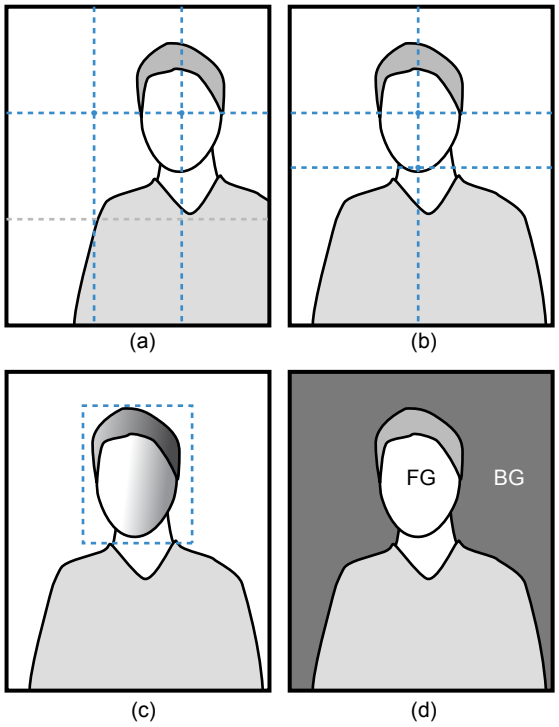


Figure 1: Composition principles used in portrait photography: (a) rule-of-thirds using top power-line points only; (b) centering on top power line or image center; (c) side-to-side contrast for defining face illumination; (c) face foreground (FG) and background (BG) contrast. ( [Chi08, Ric05, Wil])

### 2.2. Computational Evaluation of Aesthetics

The typical approach for evaluating visual aesthetics in images is to extract features and model it as a machine learning problem. Then, using a classifier, images are tagged as *aesthetic* or *non-aesthetic*; or, using a regression model, a prediction of the *aesthetic rating* is made. Below we review various methods and techniques that attempt to evaluate aesthetic images. Most methods extract many statistical local and global characteristics, and although some incorporate features which capture higher level aspects of aesthetics, these remain grouped with a large number of statistical low-level features. Also, most studies attempt to evaluate aesthetics across all image classes (e.g. landscapes, portraits, still lifes, etc.).

Datta et al. [DJLW06] are a good example of this traditional approach. They compute features based on color, light, and the rule-of-thirds and features relevant to photography such as a low depth-of-field indicator, a colorfulness measure, a shape convexity score, and a familiarity measure. In total they compute 56 different features. Classification and regression using these features is accomplished with support vector machines (SVM) and decision trees. Jiang et al. [JLC10] also use a combined regression and classification framework with 88 features based on colorfulness, contrast, symmetry, position, histogram of vanishing points, Ke's measure, multidimensional image index value, and number of faces. Dhar et al. [DOB11] use 26 features based on composition, content, and scene lighting quality. After training a SVM classifier, they showed improvements over a baseline aesthetic classifier used in Ke et al. [KTJ06]. Wong and Low [WL09] estimate salient regions using a visual saliency model which extracts multi-scaled intensity, color, and orientation features from images and finds salient locations using a neural network architecture [IKN98]. They extract 44 features: 21 global features similar to those discussed above and 23 local features based on subject and background salient regions, such as HSV, sharpness, average wavelet coefficients, saliency map, texture, mean HSV, and edge spatial distribution difference. They use these features with SVM to classify aesthetic goodness and report better performance compared to Datta et al. [DJLW06] and Ke et al. [KTJ06].

Obrador et al. [OSHO10, OSSO12] compute features that approximate traditional photography composition guidelines, such as simplicity, rule-of-thirds layout, and visual balance (using the golden mean and golden triangles). They propose a template-based feature extraction approach with 55 features. They report improvements over Luo and Tang [LT08] and Datta et al. [DJLW06]. Fedorovskaya et al. [FNH08] propose that the key to aesthetic appeal in photos is 'harmony', which they express in 16 low-level local features such as edge contrast, average lightness, and range of lightness, number of segmented regions, and scale-invariant feature transform (SIFT). Nishiyama et al. [NOSS11] also assess the aesthetic quality of photographs based on harmony, specifically colour harmony. They compute many local and global features including Moon-Spencer, chroma, red-green-blue colour vector, and hue. In total their local descriptor has 200 features. They also use a SVM classifier.

Luo et al. [LWT11] present a content-based photo quality assessment method that extracts salient regions from the image using clarity-based subject area detection [LT08], layout based surface recovery from an outdoor image [HEH07], and face detection using dynamic cascades [XZST07] combined with histograms of oriented gradients approach [DT05]. They compute 8 local features based on the clearness, colorfulness, complexity, brightness, and lighting effects and 7 global features are based on hue and scene composition. Cerosaletti and Loui [CL09] suggest that the best insight into aesthetic features is provided by dividing the images into people and non-people groups. They extract 11 features for people images and 17 features for non-people images (technical image quality, location of vanishing points, facial expressions, and location of the main subject) and perform principal component analysis and cluster analysis to group the images in similar clusters. Li et al. [LGLC10] focus on consumer photos with faces where they identify 17 features based on technical features (related to the quality of the camera equipment and the techniques used by the photographer), perceptual features (symmetry, composition, colorfulness, and consistency), and social relationship features (proximity of people) using salient face regions. They compared both classification and regression performance of this feature set against ratings generated by *Acquine*, an online machine learning based aesthetic quality prediction system for images [DW10]. Their feature set had lower residual sum-of-squares error.

The previous work we surveyed use between 26 and 200 global and local features, primarily based on statistical analysis of colour and texture. Some methods segment salient regions but the emphasis is predominantly on global features across a general space of images with little concern for aesthetics of a subclass of image types (e.g. human photos) except for Cerosaletti and Loui [CL09] and Li et al. [LGLC10]. These bottom-up approaches involve calculating many features, and then trying to correlate them with visual aesthetics. Instead, we follow a top-down approach with a very targeted image domain and a focused understanding of aesthetic criteria to isolate relevant features a-priori. Our very small set of features are primarily based on spatial and colour composition of human portraiture which we will show is actually more effective than computing a number of statistical features.

## 3. Features for Assessing Human Portraits

In the following subsections we describe the traditional set of features used in past studies and our smaller set of features based on an understanding of composition principles in portraiture. To make saliency tractable and to recognize that compositional principles are related to image content, we restrict our photographic image domain to human portraits with a single face. In our experimental framework, features from the traditional set and our proposed set are extracted from images in a portrait photo dataset (details in Section 4.1). We use different classifiers to rate each portrait as 'aesthetic' or 'non-aesthetic,' and compare this with a ground-truth aesthetic rating contained in the dataset to determine feature set performance.

### 3.1. Traditional Features

Past work has developed many different global and local features which 'should' discriminate between aesthetic and

unaesthetic images [DJLW06]. Since these were applied to general images, it follows that they will also perform well in a subset of images, in our case the restricted domain of individual photographic portraits. In this case, local features are related to the face and the global features are related to the entire image. The 66 traditional features we extracted for every portrait photograph are:

- Colorfulness
  - average Hue, Saturation, Value (HSV) of entire image, middle rule-of-thirds rectangle [DJLW06], and face [LC09]
  - average Luminance (Y) [OSSO12] and Chrominance ($C_b$ and $C_r$) [NOSS11] of entire image , middle rectangle from rule-of-thirds, and face
  - difference between global and local average Hue, Saturation and Value [WL09]
- Composition
  - distance of face mid-point from each power points (the four points where imaginary horizontal and vertical meet using the rule-of-thirds) [BSS10]
  - position of horizon, and variation from golden ratio [BSS10]
  - ratio of area of face to image [LC09, LWT11]
  - distance between center of face to any thirds-line, product of minimum of distance between center of face to 4 power points and minimum distance to any thirds-line [DOB11]
  - size and aspect ratio of image [DJLW06]
- Texture : Sharpness [WL09], Contrast [WL09], Homogeneity [OSHO10], Hough Peaks of the image [LWT11], Correlation, Energy of image and face
- Statistical : Mean, Standard Deviation, Skewness [LC09] and Kurtosis of the image and face

Our hypothesis is that this type of bottom-up approach, where a large set of features are computed and sophisticated selection mechanisms attempt to isolate the most relevant ones, is counter-intuitive. There are also problems detrimental to building a generalized classification model when using many features such as feature redundancy, over-fitting, and mutual cancellation of inverse correlated features.

### 3.2. Proposed Features

Our proposed features are different than the traditional feature set. They are focused on spatial and colour composition of human portraits which we argue is more informative, more concise, and better equipped to rank a fundamental aspect of aesthetics – image composition. As we shall see, our proposed feature set size is almost $1/9^{th}$ the size of the traditional features, yet they result in better classification performance.

#### 3.2.1. Features for Spatial Composition

Obrador et al. [OSHO10] stress the importance of composition when evaluating image aesthetics. They present a simple template-based method for computing visual composition features. Each template captures variations of a spatial composition principles: rule-of-thirds, golden mean, and golden triangle (see Figure 2). Their golden triangle principle is expressed as multiple feature templates covering rotation and symmetry permutations. Due to the multiplicity of principles and variations, the net effect of the combined template set suggests a simpler rule. Essentially a composition is penalized when not predominantly centred or in the corners. When applied to portraits, this means that as long as a face is not located along a middle edge, it will have a good feature score. We use a more selective and refined template approach to capture principles tuned to spatial composition principles of portraiture (illustrated in Figure 1):

- we use the rule-of-thirds only and prioritize power points along the top power line
- we add additional power points to reward perfectly centred portraits: one point centred on the top power line and the other centred in the template
- we compute one feature for this refined template

Our template is a two dimensional ($300 \times 300$) lookup table which returns a score given the location of the face centroid. Lighter areas are higher scores and blur locations that are close, but not exact, also have good scores (Figure 3). Face centroids located along the lighter regions have stronger spatial composition, and thus are more aesthetically appealing. In practice we use three closely related templates, one for only power-points ($f_1$), one for only power-lines ($f_2$), and the combined template ($f_3$) shown in Figure 3.

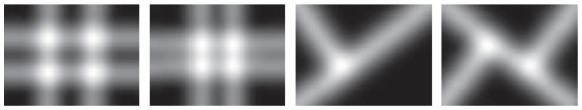

Figure 2: Spatial composition template examples from Obrador et al. [OSHO10]

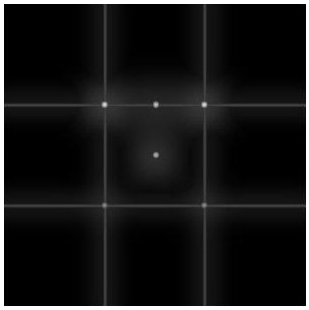

Figure 3: Spatial composition template used for lookup-table feature extraction

#### 3.2.2. Features for Highlight and Shadow Composition

In addition to spatial composition, the composition of highlights and shadows (light areas and dark areas) are important factors for aesthetic appeal. To define the shape of the face, it is preferable to illuminate the face using a soft side light, so that one side of the face is partially in shadow and the other highlighted. To focus on the individual, there should be adequate contrast between the face and the background. It is preferable to have an overall lightness in portraits, rather than an overall darkness. These qualities are captured in the following features (illustrated in Figure 1):

- face illumination ($f_4$): absolute difference between mean V (in HSV) of left and right side of face bounding box
- background contrast ($f_5$): absolute difference between mean V of face bounding box and image without face bounding box
- brightness ($f_6$): mean V of image
- size of face ($f_7$): ratio of face area to image area [LC09, LWT11]

Using the two proposed feature extraction methods, we compute only 7 features per image.

## 4. Experiments and Results

To compare the large traditional feature set (described in Section 3.1) with our small tuned feature set (described in Section 3.2), we conducted experiments using an existing data set of portrait photographs. We extracted the two feature sets from portrait photographs in the data set, and used features in each set to train five commonly used classifiers. We then compare performance of the feature sets by comparing performance of the classifiers.

### 4.1. Data Set

The human photo data set [LGLC10] consists of 500 images collected from flickr public data. The aesthetic scores of the images were collected by a survey conducted on Amazon Mechanical Turk. More than 190 "turkers" participated, with 91 ratings on more than 100 images. Each image received an aesthetic rating between 1 and 10 from a minimum of 40 people. A single aesthetic score was generated for each image by averaging all ratings it received. Since our study focuses on human portraits with a single face, we selected only 145 photos with a single person. To create a ground truth classification, we follow the method suggested by Li and Chen [LC09] where the median of all ratings is a threshold for labeling portraits as "low-quality" and "high-quality." The median for our data set is 6.9249. We label every image with an average rating less than this threshold to be aesthetically *non-pleasing* otherwise as aesthetically *pleasing*. Therefore, our data has 73 aesthetically pleasing and 72 aesthetically non-pleasing human portraits.

### 4.2. Methodology

The feature extraction scripts are developed in MATLAB (code available at `http://www.cs.uwaterloo.ca/~s255khan/code/human_face.zip`). To detect faces, we used a freely available script [Nil07] that uses successive mean quantization transform features with a split up Sparse Network of Winnows (SNoW) classifier (based on Nilsson et al. [NNC07]). This provides an axis-aligned bounding box of the salient face region. Detection was not perfect: if multiple faces were detected, the largest is used; if no face was detected (25 out of 145 images), the face bounding box was manually specified.

Five common classification algorithms were used from the Weka library [HFH*09]: K-Nearest Neighbour (for K=9), Support Vector Machine (SVM), Random Forest (RF), Classification Via Regression (CVR), and Multiboosting AdaBoost (MAB). The KNN classifier was tuned for K=9 and all other classifiers used Weka default parameter values. To avoid sampling bias, we perform 10-fold cross validation on every classifier and repeated the process 100 times, each time randomizing the order of the data set.

### 4.3. Results

Table 1 shows the results of our experiments comparing the traditional set of features (from Section 3.1) and our proposed set of features (from Section 3.2). The values in the table represent mean accuracy obtained across all runs for each classifier. It can be seen that all classifiers perform better when they are trained with our proposed features and the SVM classifier marginally outperforms all others. To ascertain if these are statistically significant differences, we use confidence intervals for two-tailed hypothesis testing under the assumption of a normal distribution. Examining the separation of 95% confidence intervals (shown as error bars in Figure 4), we see that our proposed feature set is significantly better than the traditional features with the same classifier. An additional observation from Table 1 is that the variability of results for every classifier (in terms of standard deviation of mean accuracy over 100 runs) appears lower when using our proposed feature set. With the added evidence of smaller confidence intervals for our proposed feature set, this suggests our feature set is more robust.

#### 4.3.1. Importance of Features

In order to evaluate the relative merit of the proposed features over each other, we employed three commonly used feature ranking methods from the Weka Library: Information Gain, Chi-Squared, and Probabilistic Significance as a two way function [AD05]. All of these methods showed the same feature ranking. In decreasing order of importance:

$$\underbrace{f_3, f_1, f_2}, \underbrace{f_5, f_6, f_7, f_4}$$

Table 1: Classification accuracy on applying both feature extraction methods.

| *%age Accuracy (Standard Deviation)* | | |
|---|---|---|
| ***Classifier*** | **Traditional features** | **Proposed features** |
| SVM | 57.66 (0.023) | **61.10** (0.017) |
| RF | 57.75 (0.022) | 59.79 (0.018) |
| KNN | 56.20 (0.018) | 59.33 (0.014) |
| CVR | 57.28 (0.016) | 59.92 (0.012) |
| MAB | 58.08 (0.014) | 59.14 (0.012) |

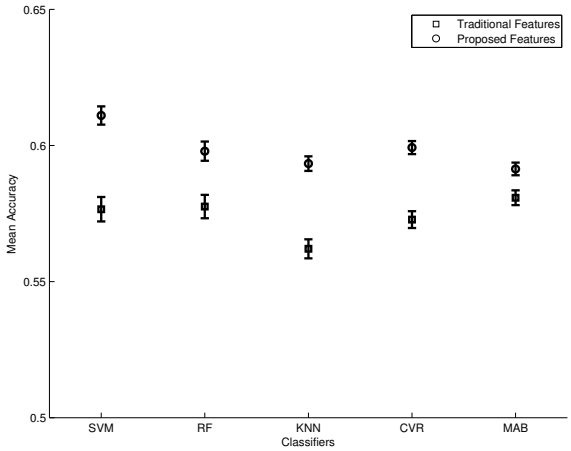


Figure 4: Mean classifier accuracy of two features set (95% CI error bars, note y-axis ranges between 0.5 and 0.65 only)

The top three features are the template-based features for spatial composition, followed by four features for highlight and shadow composition. This analysis suggests that spatial composition is more important than highlight and shadow composition when determining aesthetics.

#### 4.3.2. SVM Parameter Tuning

The SVM has a complexity parameter $C$, which controls the softness of the class margins or the number of data objects that are used as 'support vectors' to draw the linear separation boundary in transformed higher dimensional space. In the main experiment we used the default value of $C = 1$. To tune the value of $C$, we evaluated accuracy using a 10-fold cross validation for $C = \{1, 2, ..., 10\}$. We found that we could increase the accuracy to 63.51% with $C = 3$.

### 4.4. Discussion

Our results show that the proposed reduced feature set encodes more information about the aesthetic quality of human portraits, even though it is almost $1/9^{th}$ the size of the traditional feature set. This result provides evidence that features based on top-down composition principles contain greater discriminatory power and are more effective in the evaluation of visual aesthetics. Using the traditional approach of searching for a correlation between a large set of global statistical features and aesthetic classification is less effective in spite of requiring more work and computational power. This verifies the intuition that understanding aesthetics to develop 'relevant features' can help in building better classifiers to model visual aesthetics in human portraits. It is important to recognize a potential limitation of our approach, specifically our focus on portraits of individuals. Individual portraits are not necessarily the most common type of portrait, much less the most common type of photograph. However, they certainly exist in some number, and moreover, it serves to support our argument that tailoring a small set of features to a specific image class has computational, conceptual, and performance advantages. With even better salient information – like the age and gender of the person, what objects are in the background, or the type of event such as birthday, wedding, etc. – our compositional principles could be further refined and additional principles conditionally introduced. Regardless, our work is a step closer to a smart phone application for casual portrait photographers, where immediate aesthetic assessments are provided to encourage re-taking photos for better composition, or assisting with decisions for deleting, sharing, and posting portrait photos.

#### 4.4.1. Case Study

To illustrate how our proposed features classify aesthetic and non-aesthetic photographic portraits, we ran a case study experiment. We trained the tuned SVM classifier (with $C = 3$) on 130 randomly selected portraits from the data set (90%) and classified the remaining 15 human portraits (10 aesthetic and 5 non-aesthetic photos). The classifier correctly identified 7 aesthetic photos (TP) and 4 non-aesthetic photos (TN). Three aesthetic photos were wrongly identified as non-aesthetic (FN) and one non-aesthetic photo is wrongly identified as aesthetic (FP) (examples are shown in 5). The TP and TN portraits most closely adhere to our notion of compositional principles for portraits.

## 5. Conclusions and Future Work

Past efforts to assess aesthetic quality in photographs have primarily used a bottom-up approach where a large number of local and global features are extracted. We believe this may be symptomatic of difficult saliency detection and the goal of classifying *all types* of images. In contrast, we use a top-down approach using only 7 features built on an understanding of compositional principles tailored to the constrained image domain of photographic portraits — where saliency detection is more tractable. Compared to the large feature sets traditionally used, our method is up to 6% more accurate in spite of using $1/9^{th}$ the number of features.

Although we focus on photographic portraiture as the constrained image scenario, we believe this can be extended to

(TP) correctly classified as aesthestic

(TN) correctly classified as un-aesthestic

(FN) incorrectly classified as un-aesthestic

(FP) incorrectly classified as aesthestic

Figure 5: Example images from case study classification. Faces in the images are blurred and taken from the work of Li et al. [LGLC10]

other types of photographs. A natural extension is to photographs of small and large groups of people where multiple faces introduce more complex spatial compositions. Another practical area are photographs of items offered for sale on auction sites, classified advertising sites, and retail databases. To finely tune compositional principles, this could be constrained by product type: houses, automobiles, pottery, furniture, etc. We believe that Obrador et al. [OSHO10] look-up table approach is essential for the features encoding spatial composition, and perhaps other aspects having an influence on aesthetics as well.

## References


[AD05] AHMAD A., DEY L.: A feature selection technique for classificatory analysis. *Pattern Recognition Letters 26*, 1 (2005), 43–56. 5

[BSS10] BHATTACHARYA S., SUKTHANKAR R., SHAH M.: A framework for photo-quality assessment and enhancement based on visual aesthetics. In *ACM Multimedia* (2010), Bimbo A. D., Chang S.-F., Smeulders A. W. M., (Eds.), ACM, pp. 271–280. 1, 4

[Chi08] CHILD J.: *Studio Photography: Essential Skills*, 4th ed. Focal Press, 2008. 1, 2

[CL09] CEROSALETTI C. D., LOUI A. C.: Measuring the perceived aesthetic quality of photographic images. In *First International Workshop on Quality of Multimedia Experience (QoMEX)* (2009), pp. 47–52. 1, 3

[Dic08] DICKSON C. N.: 6 Tips for Perfect Composition in Portrait Photography. `http://www.digital-photography-school.com/6-tips-for-perfect-composition-in-portrait-photography`, 2008. Online Accessed 12$^{th}$ Mar 2012. 2

[DJLW06] DATTA R., JOSHI D., LI J., WANG J. Z.: Studying aesthetics in photographic images using a computational approach. In *ECCV (3)* (2006), Leonardis A., Bischof H., Pinz A., (Eds.), vol. 3953 of *Lecture Notes in Computer Science*, Springer, pp. 288–301. 1, 3, 4

[DOB11] DHAR S., ORDONEZ V., BERG T. L.: High level describable attributes for predicting aesthetics and interestingness. In *CVPR* (2011), pp. 1657–1664. 1, 3, 4

[DT05] DALAL N., TRIGGS B.: Histograms of oriented gradients for human detection. In *CVPR (1)* (2005), IEEE Computer Society, pp. 886–893. 3

[DW10] DATTA R., WANG J. Z.: Acquine: aesthetic quality inference engine - real-time automatic rating of photo aesthetics. In *Multimedia Information Retrieval* (2010), Wang J. Z., Boujemaa N., Ramirez N. O., Natsev A., (Eds.), ACM, pp. 421–424. 3

[FNH08] FEDOROVSKAYA E. A., NEUSTAEDTER C., HAO W.: Image harmony for consumer images. In *ICIP* (2008), IEEE, pp. 121–124. 3

[HEH07] HOIEM D., EFROS A. A., HEBERT M.: Recovering surface layout from an image. *International Journal of Computer Vision 75*, 1 (2007), 151–172. 3

[HFH*09] HALL M., FRANK E., HOLMES G., PFAHRINGER B., REUTEMANN P., WITTEN I. H.: The WEKA data mining software: an update. *SIGKDD Explorations 11* (2009), 10–18. 5

[IKN98] ITTI L., KOCH C., NIEBUR E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell. 20*, 11 (1998), 1254–1259. 3

[JLC10] JIANG W., LOUI A. C., CEROSALETTI C. D.: Automatic aesthetic value assessment in photographic images. In *ICME* (2010), IEEE, pp. 920–925. 3

[Kra05] KRAGES B. P.: *Photography: The Art of Composition*. Allworth Press, New York, N.Y., 2005. 2

[KTJ06] KE Y., TANG X., JING F.: The design of high-level features for photo quality assessment. In *CVPR (1)* (2006), IEEE Computer Society, pp. 419–426. 1, 3

[LC09] LI C., CHEN T.: Aesthetic visual quality assessment of paintings. *IEEE Journal of Selected Topics in Signal Processing 3*, 2 (2009), 236–252. 4, 5

[LGLC10] LI C., GALLAGHER A. C., LOUI A. C., CHEN T.: Aesthetic quality assessment of consumer photos with faces. In *ICIP* (2010), pp. 3221–3224. 2, 3, 5, 7

[LT08] LUO Y., TANG X.: Photo and video quality evaluation: Focusing on the subject. In *ECCV (3)* (2008), Forsyth D. A., Torr P. H. S., Zisserman A., (Eds.), vol. 5304 of *Lecture Notes in Computer Science*, Springer, pp. 386–399. 3

[LWT11] LUO W., WANG X., TANG X.: Content-based photo quality assessment. In *ICCV* (2011). 1, 3, 4, 5

[Nil07] NILSSON M.: Face Detection in Matlab. `http://www.mathworks.com/matlabcentral/fileexchange/13701-face-detection-in-matlab`, 2007. Online Accessed $12^{th}$ Mar 2012. 5

[NNC07] NILSSON M., NORDBERG J., CLAESSON I.: Face Detection using Local SMQT Features and Split Up SNoW Classifier. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (to 2007), vol. « to be updated ». 5

[NOSS11] NISHIYAMA M., OKABE T., SATO I., SATO Y.: Aesthetic quality classification of photographs based on color harmony. In *CVPR* (2011), pp. 33–40. 1, 3, 4

[OSHO10] OBRADOR P., SCHMIDT-HACKENBERG L., OLIVER N.: The role of image composition in image aesthetics. In *ICIP* (2010), pp. 3185–3188. 2, 3, 4, 7

[OSSO12] OBRADOR P., SAAD M. A., SURYANARAYAN P., OLIVER N.: Towards category-based aesthetic models of photographs. In *MMM* (2012), Schoeffmann K., Mérialdo B., Hauptmann A. G., Ngo C.-W., Andreopoulos Y., Breiteneder C., (Eds.), vol. 7131 of *Lecture Notes in Computer Science*, Springer, pp. 63–76. 3, 4

[Ric05] RICE P.: *Professional Techniques for Black and White Digital Photography*. Amherst Media, Inc., 2005. 1, 2

[Wil] WILLIAMS P.: 10 Top Photography Composition Rules. `http://www.photographymad.com/pages/view/10-top-photography-composition-rules`. Online Accessed $12^{th}$ Mar 2012. 2

[WL09] WONG L.-K., LOW K.-L.: Saliency-enhanced image aesthetics class prediction. In *ICIP* (2009), IEEE, pp. 997–1000. 3, 4

[XZST07] XIAO R., ZHU H., SUN H., TANG X.: Dynamic cascades for face detection. In *ICCV* (2007), IEEE, pp. 1–8. 3