The Beauty of Capturing Faces: Rating the Quality of Digital Portraits

Miriam Redi[1], Nikhil Rasiwasia[2], Gaurav Aggarwal[2], Alejandro Jaimes[3]

[1] Yahoo Labs, Barcelona, Spain [2] Yahoo Labs, Bangalore, India [3] Yahoo Labs, New York, USA

Abstract— Digital portrait photographs are everywhere, and while the number of face pictures keeps growing, not much work has been done to on automatic portrait beauty assessment. In this paper, we design a specific framework to automatically evaluate the beauty of digital portraits. To this end, we procure a large dataset of face images annotated not only with aesthetic scores but also with information about the traits of the subject portrayed. We design a set of visual features based on portrait photography literature, and extensively analyze their relation with portrait beauty, exposing interesting findings about what makes a portrait beautiful. We find that the beauty of a portrait is linked to its artistic value, and independent from age, race and gender of the subject. We also show that a classifier trained with our features to separate beautiful portraits from non-beautiful portraits outperforms generic aesthetic classifiers.

Fig. 1. (a) Distribution of images per demographic category and aesthetic scores (b) Characteristics extracted from the portrayed subjects

I. INTRODUCTION

Portraits make up a large percentage of the photos on the web nowadays. "Selfies" have become a phenomenon, and recent studies [1] show that images with faces are more popular (+38% "likes" on Instagram) than other pictures in online social networks. Portraits are also used in web user profiles, in news articles, to represent celebrities and public figures, and they are an essential part of all kinds of IDs.

Given the huge volume of digital portraits, their broad usage, and their importance for people identification, surfacing the best digital portraits in terms of photographic quality is of crucial importance. A system able to automatically score the aesthetic value of portraits could be used to select good images for a variety of applications such as journalism, photo sharing websites, web search, PhotoBoosts, and many others.

Shooting photos of people is not a trivial task: human faces convey emotions, stories, lifestyles, and a good photographer needs to be able to capture their essence and personality. As a matter of fact, portrait photography is a stand-alone branch of photography literature, with its own rules and compositional techniques, and tons of dedicated books [2], [3], [4]. Systems that automatically rate the quality of digital portraits should be therefore specifically designed for face photos, unlike traditional visual aesthetics works [5], [6], based on general photographic rules .

In spite of its importance, there has been little work in the research community to specifically address computational aesthetics of portraits. Preliminary works [7], [8] leave out many of the aspects that are specific to portraits (e.g., illumination, landmark representation, affective properties, etc.), and have experimented only with small datasets (less than 500 images).

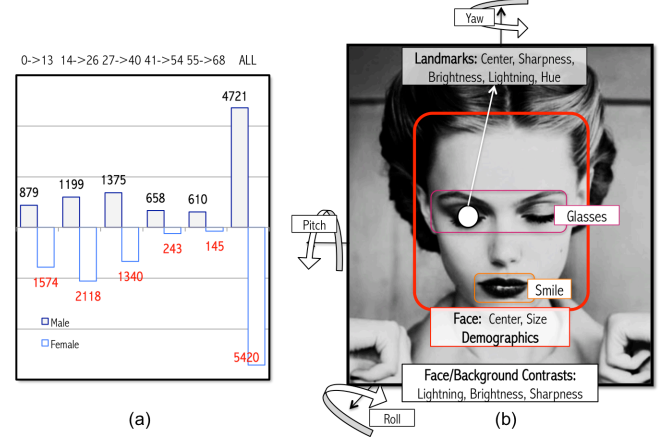In this paper, we try to fill this void and introduce a new framework to automatically evaluate portrait aesthetics[1]. To do so, we design visual features to describe image quality and portrait-specific properties and present a large-scale analysis of a data set of over 10,000 portraits. In addition, we build predictive models that are able to determine the aesthetic score of digital portraits. Moreover, with such large scale study, we provide an analysis of what makes a portrait beautiful from a computational perspective. To our knowledge, this represents the first attempt in literature to understand the relevancy of features for portrait aesthetics.

Our main contributions can be summarized as follows:
(1) Dataset: we build a large dataset of portraits annotated with physical characteristics (determined using facial analysis) by sampling the AVA [9] images.
(2) Features: we introduce new features to describe portrait composition, quality, illumination, memorability, emotions, and originality.
(3) Feature analysis: we perform analyses on a set of over 10,000 portraits and report observations. We find that race, gender, and age are largely uncorrelated with photographic beauty, but aesthetic score is related to sharpness of facial landmarks, image contrast, exposure, homogeneity, illumination pattern, uniqueness, and originality.
(4) Aesthetic Prediction: we develop predictive models to classify portraits as aesthetically beautiful or not.

In Sec. II, we describe related work, and explain our portrait dataset in Sec. III. Sec. IV presents the visual features and analyze their relations with portrait beauty in Sec. V, then present our classification experiments in Sec. VI.

---

[1]The aim of this work is to estimate the photographic quality of the *representation* of the person, independent from the beauty of the subject represented.

II. Related Work

Our work relates to research that applies image analysis techniques to detect the visual presence of non-semantic, fuzzy concepts such as memorability [10], emotions [11], [12], interestingness [13], [14], [15], privacy [16], and beauty [5]. In particular, this paper follows previous work on computational aesthetics [5], [9], that explores the discriminative ability of visual features to automatically assess the beauty of images and videos. Pioneers in this field are Datta et al. and Ke et al., [5], [6], who built an aesthetic classification framework for images based on features inspired by photographic theory. In subsequent years, such works were improved by designing more discriminative features [17], [14], proving the effectiveness of generic features [18], [9] and building more effective learning frameworks [19]. Similar frameworks were applied to automatic image composition and enhancement by Bhattacharya et al. [20].

While these existing computational aesthetic works build general frameworks for photographs of any semantic category, we focus on a specific type of images, namely portraits, whose compositional and aesthetic criteria constitute a separate subject of study in the photographic literature [3], [4], [2], and therefore need a separate computational framework for aesthetic assessment. This aspect is also proven by our experiments: we show that our portrait-specific aesthetic framework performs much better than a general classifier for portrait aesthetic assessment.

A few works [9], [21], [14] perform topic-based aesthetic classification. They build category-specific subsets of images by sampling aesthetic databases according to given image tags ("city", "nature", but also "humans" or "portraits"), and then use general compositional features to build topic-specific models. The framework in this paper differs from those works for two reasons. (1) *We build a rich large-scale portrait aesthetic database*. A dataset based on tag-based sampling as in [9], [21], [14], could ignore many face images without tags while including images with noisy tags (as shown in Section III). In this paper, we adopt a *content-aware sampling strategy* based on detailed face analysis. We reduce a large scale aesthetic dataset [9] to a subset of more than 10000 face images annotated with information about the portrayed subject, useful for both analysis and feature extraction. (2) *We build portrait-specific aesthetic visual features*. The works in [9], [21], [14] use traditional aesthetic features designed for a general case, and apply them to the topic-specific contexts. In our work, we design face-specific aesthetic features inspired by photographic literature, together with non-face features that describe crucial aspects of photographic portraiture, such as illumination, sharpness, manipulation detection, image quality, emotion and memorability. Moreover, we show their combined effectiveness for aesthetic assessment of face photographs compared to traditional aesthetic features.

There are a few recent works that attempt to design portrait specific datasets and features. For example, Li et al.[8] use face expression, face pose and face position features to estimate the aesthetic value of the images in a dataset of 500 face images annotated by micro-workers. This work was improved by the work in [22], that uses hand-crafted features together with low-level generic features, and by Khan et al. [7] using spatial composition rules specifically tailored for portrait photography, together with specific background contrast features and face brightness and size features. These works represent a first attempt towards portrait aesthetic classification. However, one major weak point of such works is that they rely on small datasets (¡500 images), thus making the results less generalizable for large datasets like the one we consider. Moreover, despite their focus on face analysis, the features proposed by those works miss many important aspects of portrait photography such as illumination, demographics, face landmark properties, affective dimension, semantics and post-processing. In our work, we use features that are able to capture these aspects, and prove their effectiveness by showing that they outperform the features in [7] when used in an aesthetic classification framework on the dataset used by Khan et al. [7]. Moreover, in this paper, we perform for the first time a deep analysis of the importance of each feature and each group of features for face photo aesthetics, giving interesting and probably unexpected insights about what makes a portrait beautiful.

III. Large Scale Portrait Dataset

In order to create a large scale corpus of face images annotated with beauty scores, we resort to the largest aesthetic database available in the literature, i.e. the AVA dataset [9], created from the photo challenge website dpchallenge.com, that contains more than $250,000$ images annotated with an aesthetic score, a challenge title, and semantic textual tags.

AVA is a unique, rich dataset for visual aesthetics, and therefore a reliable source of data for our purposes. However, AVA images contain very diverse subjects other than faces. Moreover, for analysis and classification purposes, we want to collect not only a reliable subset of portrait images, but also some rich information about the portrayed subject and its representation. With this in mind, we design a *content-aware sampling strategy* on the AVA dataset, based on both metadata-based filtering and face analysis:

(1) **Enhanched metadata-based filtering.** First, we select from the AVA database not only the images tagged as "Portrait" but also all the images whose challenge title contains the words 'Portrait', 'Portraiture' or 'Portraits'. (e.g. *Portrait Of The Elderly*). A total of 21,719 images are collected at this stage.

(2) **Face detection-based filtering.** We use Face++ [23] to filter the images collected after metadata-based filtering. We obtain a subset of 10,141 images for which Face++ detected the presence of one or more faces (in case of multiple faces, we retain the information about the largest one only).

(3) **Subject properties.** We compute though Face++ basic information about the subject, such as position, orientation, demographics (race, gender, age), coordinates of facial landmarks (eyes, nose and mouth in relative coordinates),

presence of smile, presence of glasses, etc. (for a complete list of features see Table I).

For each of the resulting images, we assign the average aesthetic score (in a 1-10 range) according to the votes provided by the AVA dataset. Figure 1 shows the composition of our dataset, highlighting the distribution, based on gender and other properties estimated by the Face++ detector. About 53% of the subjects are classified as female, and 1/3 of the image corpus shows subjects between 14 and 26 years of age (Fig. 1 (a)). Similar to the AVA dataset, the vast majority of the aesthetic scores lies between 4 and 6, with a peak around the mean, which stands at 5.5.

IV. FEATURES FOR PORTRAIT AESTHETIC ASSESSMENT

Visually stunning portrait photographs are often the result of an artistic process that might not strictly follow general rules of composition, or fulfill basic quality requirements. However, photographic portraiture literature [2], [3], [4] suggests that following some specific photographic principles can help making digital portraits more attractive, ensuring visual appeal and expressiveness. Among the various tips for good portraiture available in literature, we identified 5 main photographic dimensions, namely:

Compositional Rules: arrangement of lines, objects, lights and color, widely used in visual aesthetic literature [5], [21].

Scene Semantics: where has the photo been shot? and which objects co-exist with the subject in the scene?

Portrait-Specific Features: information about the subject (aspect, soft biometrics, demographics) and its representation (sharpness, illumination, etc.)

Basic Quality Metrics: principles that ensure the correct perception of the signal, without distorting the scene represented. Rarely used in computational aesthetics, they can be fundamental for high-quality portraiture [3].

Fuzzy Properties: portrait photographic beauty is related to non-objective properties such as emotions or uniqueness, which are unquantifiable with low level features.

In this work, we design 5 groups of features that aim at describing various aspects of each of these dimensions using computer vision techniques.

A. Compositional Rules

As highlighted in many previous works [5], [32], [20], the visual attractiveness of a picture is strongly influenced by the arrangement of objects in the image, their lighting, their colors, their perceptibility.

Similar compositional rules apply to portraits photography. However, since portraits generally focus on a single subject whose essence needs to be captured in the shot, two compositional aspects need particular consideration: lighting and sharpness. The correct illumination of the scene and the detailed representation of the subject ensures both perceptibility and expressiveness. Given these observations, we design a set of new features that capture essential properties of image lighting and sharpness, and collect a set of existing features for image composition analysis.

Lighting Features
The lighting setup is crucial to determine the essence of the portrait. In previous works [5], [7], [21], scene lighting is described using features based on overall image brightness. However, as proved by our results, the raw brightness channel information might not be enough to capture portrait lighting patterns.

We therefore design a new lighting feature to expose **Lighting Patterns** based on an illumination compensation algorithm originally created for face recognition [33]. Such method considers an image I as a product $I = R(I) \cdot L(I)$, where $R(I)$ is the 'reflectance' of the image and $L(I)$ is its "illuminance'" i.e. the perceived lighting distribution.

In order to infer the lighting pattern of an image, we proceed as follows. For each image, we calculate $L(I)$ and create an illuminance vector $V(I)$ by averaging its illuminance $L(I)$ over local windows (25x25 subdivision). Applying k-means clustering on the illuminance vectors of a set of training images, we group the illuminance vectors into 5 Lighting Patterns representing the most common lighting setups in our dataset (See Fig.2). For a new image J, we assign its corresponding lighting pattern by looking at the closest cluster to its illuminance vector $V(J)$, and retain the cluster number as the *Lighting Pattern Feature*.

Sharpness Features
The recognizability and sharpness of the subject is a basic requirement for good portraiture. To analyze the amount of sharpness in the image, we design two new features:

Overall Sharpness: Subject movements or camera defocus can affect the overall image sharpness, introducing disturbing blur in particular image regions. We compute the sharpness of a picture by calculating the strength of the edges after applying horizontal and vertical Sobel masks on the image, according to the Tenengrad method (as explained in [34]).

Camera Shake: sometimes camera movements can create an overall blurriness in the image. In order to estimate this particular type of blur, we compute the ratio between the number of pixels detected to be affected by camera shake and the total number of pixels, according to the camera motion estimation algorithm of Chakrabarti et al. [24].

Traditional Compoisitional Features
We collect here a set of features from state-of-the art works that model compositional photographic rules using a computational approach.

Color Features. In order to capture color patterns and their relation with portrait aesthetics, we compute the following features extracted from literature: *Color names* [11], *Hue, Saturation, Brightness (HSV)* [11], [5] , the *Pleasure, Arousal, Dominance* metrics [11], the *Itten Color Histograms* [11], and the corresponding *Itten Color Contrasts*: [11] Moreover, we compute 2 contrast metrics: *Contrast (Michelson)* [25], and a traditional *Contrast* measure computed as the ratio between the difference of max-min values of the Y channel and the Y average.

Spatial Arrangement Features. The distribution of textures, lines and object in the image space is an important cue

Feature	Dim	Description	References
Compositional Features			
Lightning Patterns	5	Lightning pattern according to the image illuminance	new
Overall Sharpness	1	Sum of the image pixels after applying Sobel masks	new
Camera Shake	1	Ratio between 'moving' pixels identified by the method in [24] and image size	new
Color Names	9	Number of pixels that belong to given color clusters such as *black, blue, green, flesh, magenta, purple*	[11]
HSV average	6	Average *Hue, Saturation, Brightness* of the *whole* image and in the *inner quadrant*	[11], [5]
Pleasure, Arousal, Dominance	3	Affective dimensions computed by linearly combining HSV values	[11]
Itten Color Histograms	20	Histograms of H, S and V values quantized over 12, 3, and 5 bins	[11]
Itten Color Contrasts	3	Standard deviation of the Itten Color Histograms distributions	[11]
Contrast (Michelson)	1	Ratio between the sum of max and min luminance values and their difference	[25]
Contrast	1	Ratio between the sum of max and min luminance values and the average luminance	new
Symmetry (Edge)	1	Distance between edge histograms on left and right halves of the image	[13]
Symmetry (HOG)	1	Difference between HOG features on left and right halves of the image	new
Number of Circles	1	Computed using Hough transform	new
Rule of Thirds	9	Based on saliency distribution of the 9 image quadrants resulting after a 3×3 division of the image	new
GLCM Properties	4	*Entropy, Energy, Homogeneity, Contrast* of the GLCM matrix	[11]
Image Order	2	Order values obtained through Kologomorov *Complexity* and Shannon's Entropy	[26], [13]
Level of Detail	1	Number of regions after Watershed segmentation	[11]
Semantics			
Object Bank Features	189	Object Bank image representation	[27]
Basic Quality Metrics			
Noise	1	Distance between original image and image denoised with the algorithm from [28]	new
Contrast Quality	1	Negative distance between original image and image with normalized contrast	new
Exposure Quality	1	Negative absolute value of the luminance histogram skewness	new
JPEG Quality	1	Computed with the no-reference quality estimation algorithm in [29]	[29]
Image Manipulations	2	Amount of *Splicing* and *Median Filtering* applied to the image	new
Portrait-Specific Features			
Face Position [23]	4	*X, Y* in relative coordinates, plus relative *Width* and *Height*	[23]
Face Orientation [23]	3	*Yaw, Pitch* and *Roll* angle of the head	[23]
Demographics [23]	6	Race (*White, Black, Asian*), *Age* (in years) and *Gender*	[23]
Landmark Coordinates [23]	8	*Right/Left* Eye, *Nose* and *Mouth* position in relative coordinates	[23]
Similing Expression [23]	1	Estimates wether the subject is smiling or not	[23]
Other Face Properties [23]	3	Presence of *Glasses* (none, sunglasses, normal glasses)	[23]
Landmark Statistics	12	*Hue* and *Brightness* of Right/Left Eye, Nose and Mouth	new
Landmark Sharpness	4	Sharpness of *Right/Left* Eye, *Nose* and *Mouth* using gradient magnitude	new
Face/Background Contrasts:	3	Contrast between face region and background in terms of *Lightning, Sharpness* and *Brightness*	new
Fuzzy Properties			
Emotion	1	Estimates the positive/negative traits of the emotions that the image arouses using compositional features and affective image datasets [30], [31], [11]	new
Originality	1	Estimates the image originality based on a classifier trained on the Photo.net dataset [5]	new
Memorability	1	Estimates the image memorability based on a classifier trained on the memorability dataset [10]	new
Uniqueness [13]	1	Based on the image spectrum	[13]

TABLE I
VISUAL FEATURES FOR PORTRAIT AESTHETIC MODELING

for aesthetic and affective image analysis, as proved in [5], [11], [32], [14]. To analyze spatial layout of objects and shapes in the scene, we compute first two symmetry descriptors, namely *Symmetry (Edges)*[14], and *Symmetry (HOG)*, for which we retain the difference between the HOG [35] descriptors from left half of the image, and from the flipped right half. Moreover, we compute 2 new features that describe shapes and their distribution, namely the *Number of Circles*, and the *Rule of Thirds*, that, unlike previous works [5], [20], determines the rule of thirds by computing the amount of spectral saliency [36] in the 9 quadrants resulting from a 3x3 division of the image.

Texture Features. Textural features can help analyzing the overall smoothness, order and entropy of the image. We analyze image homogeneity by computing the *GLCM properties* [11], the *Image Order* [13], and the *Level of Detail* [11].

B. Semantics and Scene Content

As proved by various works in visual aesthetics [13], [21], [32], the content of the scene and the types of objects placed in the picture substantially influence the aesthetic assessment of pictures. In particular, in the portraiture context, it is important to analyze the setting where the photo has been shot, i.e. objects, scenery and overall harmony of subject with

the scene. In order to estimate these properties, we compute an adapted version of the **Object bank features** [27] that retains the maximum probability of a pixel in the image to be part of one of the 208 objects in the Object Bank.

C. Basic Quality Metrics

In general, visually appealing portraits are also high-qiuality photographs, i.e. images where the degradation due to image registration or post-processing is not highly perceivable. In order to deeply analyze this dimension, we design some rules to determine the perceived image degradation by looking at simple image metrics, independent of the composition, the content, or its artistic value, namely:

Noise: we compute the amount of camera noise by applying an image denoising algorithm [28], and then computing the distance between the denoised image and the original one.

Contrast Quality: well-contrasted images, i.e. images where the contrast level allows to distinguish the picture shapes without introducing disturbing over-saturated regions, can be recognized by the uniform distribution of the intensities on the image histogram. We therefore compute the quality of the contrast by negating [2] of the distance between the original

---

[2]We take the negative of the distance in order to have higher values of this features for higher contrast quality
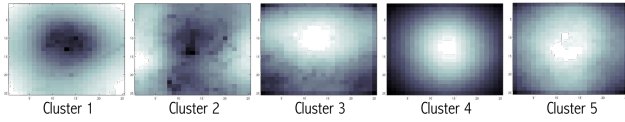
Fig. 2. Illuminance Distribution of the 5 Lightning Patterns

image and its contrast-equalized version.

Exposure Quality: the luminance histogram of an overexposed image is skewed towards the right part, while for an underexposed image it is skewed towards the left side. In order to capture this behavior, we convert the image to the YCbCr space, we compute the skewness of the Y channel histogram over 255 bins. When the skewness is close to zero, the exposure is correct, when below or above zero, the image is under or over exposed. We negate the absolute value of the skewness as exposure balance metric.

JPEG Quality: when too strong, JPEG compression can cause disturbing effects such as blockiness or block smoothness. We implement the objective quality measure for JPEG images proposed by [29] and retain the JPEG quality score output by the algorithm.

Image Manipulations: more and more, digital pictures are post-processed after the shooting using editing tools. In order to understand the amount of post-processing applied on the image, we design 2 new quality metrics, inspired by blind image forensics techniques. First, we design a feature to compute the amount of *Splicing Manipulation*: we retain the output of an SVM classifier trained with Markov Features [37] computed on a training set of images annotated as spliced/not spliced from the CASIA dataset [38] (85% accuracy on this set). Next, we build a feature to compute the amount of *Median Filtering Manipulation*, using the algorithm of Yuan et al. [39].

D. Portrait-Specific Features

In photographic portraiture, lot of effort should be spent on understanding the subject and its correct representation. Photographic portrait theory [3] particularly stresses the importance of the focus, sharpness, lighting and position of the face landmarks (eyes,nose,mouth).

In order to describe the properties of the subject and its representation, we retain as candidate features all the values extracted automatically by the Face++ api, and we build on top of such values a set of features to deeply describe the face and landmark properties. Overall, the set of Face/Subject features is as follows:

Face++ description: *Face Position*, namely x, y relative co-ordinates, plus relative width and height , *Face Orientation*, i.e., yaw, pitch and roll angle of the head, *Demographics* like Race (white, black, asian), Age (in years) and Gender, *Landmark Coordinates*, namely Right/Left Eye, Nose and Mouth position in relative coordinates, *Subject Expression* , that estimates wether the subject is smiling or not, and *Other Face Properties* such as presence of glasses (none, sunglasses, normal glasses).

Landmark Sharpness for each landmark, we simply compute its sharpness by averaging the gradient magnitude over the landmark region.

Landmark Staitstics: for each landmark, we extract its average Hue and Brightness

Face/Background Contrasts: similar to the background contrast feature in [7], we analyze here the compositional differences between face region and background region. However, while Khan et al. [7] simply retain the ratio between face region brightness and image brightness, we perform here a deeper analysis. We consider face (F) and background (B) as two separate sub-images. We then compute the *Lighting Contrast* as the ratio between the average Lightning (see Sec. IV-A) of F and the average Lightning of B, the F/B *Sharpness Contrast* (Sharpness is computed computed as for the Landmark Properties), and, similarly, the *Brightness Contrast*.

E. Fuzzy Properties

Some artistic traits of photographs cannot be directly captured by low-level features: many times, photographic beauty is related to feelings vehiculated by the image, which not even words can describe. In our work, we try to model some of those 'fuzzy' properties using a computational approach, by re-using existing work on image memorability, originality and affective analysis.

Emotion, is the emotion aroused by the image positive or negative? We address this question by training an emotion classifier (SVM, 75% accuracy) with traditional Compositional Features, using as a groundtruth a mixture of 3 affective dataset [31], [30], [11]. We binarize the annotation in order to reflect the positive/negative trait of the emotion shown. For each image , we retain the emotion score predicted by such classifier as the image emotion feature.

Originality of the image composition is computed by retaining the output of an originality classifier trained with Compositional Features and the Photo.net database from [5] (Support Vector Regression (SVR), 4,7% MSE).

Memorability of the image content. We compute this by retaining the output of a memorability classifier trained with the Saliency Moments Features[40], and the memorability database of Isola et al. [10] (SVR, 2% MSE).

Uniqueness: as in [13], we estimate the photo uniqueness as the euclidean distance between the average spectrum of the images in a database and the spectrum of each image.

V. WHAT MAKES A PORTRAIT BEAUTIFUL?

Among all the features in Section IV, which of them is more discriminative to identify beautiful portraits in a computational framework? In this Section we explore the relations between the visual features extracted and portrait aesthetic scores, by first analyzing the importance of each feature group described in Sec IV, and by then looking at the relevance of each single feature within dimensions defined.

A. Feature Groups for Portrait Aesthetics

To measure the significance of the five feature sets, we perform regression analysis using LASSO [41] for the different groups of features (i.e. *Compositional Features*). Once the

regression parameter vector is learned, we use compute the Spearman correlation between the predicted scores and the original aesthetic scores. This gives us a multidimensional correlation metric that indicates the relevance of feature group for portrait aesthetic assessment. We split the data into 5 random partitions, using one of the partitions as the test set and the rest as training, and learn regression coefficients to predict the aesthetic scores on the test set using the different groups of features.

As shown in Fig. 3(b) all the groups of features correlate positively with aesthetic scores. As expected, given the importance face of representation for portraiture, the *Portrait-Specific Features* correlate the most among all the groups of features proposed, with a correlation of 0.330 ± 0.029. Despite its rich semantic analysis, and the proved effectiveness for scene analysis [27], the ObjectBank *Semantic Features*, with its 190 feature detectors, are not as predictive, achieving a correlation score of 0.211 ± 0.022 in contrast to compositional features which achieve a correlation score of 0.290 ± 0.029. In comparison to these large feature sets, smaller sets of features such as *Basic Quality* and *Fuzzy Properties* with 6 and 4 dimensions respectively achieve a much lower correlation score for portrait aesthetics assessment as a whole, despite the importance of single features within the groups.

In order to calculate the combined predictive power of the whole set of features proposed, we perform similar regression analysis on all features together , i.e. without logical grouping, and look at the behavior of the algorithm as more and more features are taken into account. Figure 3(d) shows a plot of the Spearman correlation of the feature set as a function of the number of features used and chosen by LASSO. Using one single feature (*Right_Eye_Sharpness*), the Spearman correlation between predicted aesthetic scores and original aesthetic scores is 0.252 ± 0.018. The best correlation score of 0.398 ± 0.027 is obtained taking into account all 300 features. However, adding more than 60 features shows diminishing returns. The correlation with 60 features stand at 0.37. The smallest mean square error achieved on the test set stands at 0.430 ± 0.008.

Table II reports the weight of the features ranked by when they are first picked by LASSO. Also reported are the feature category and weights. Notice how all the feature groups appear in the top-10 features, thus confirming the importance of each dimension we consider for portrait aesthetic evaluation, with a predominance of face features. We can also spot some first insights about the importance of single features: crucial for aesthetic prediction are landmark sharpness (*Right_Eye* and *Left_Eye*), the Exposure Quality, and the high discriminative ability of the Fuzzy Properties *Uniqueness*.

B. Single Features for Portrait Aesthetics

To analyze in a more detailed manner which features correlate most with beautiful portraits, we partition the dataset into 5 subsets, as in Sec. V-A and average the

Rank	Feature Name	Feature Group	Weight
1	Left_Eye_Sharpness	Portrait Features	**0.061894**
2	Right Eye_Sharpness	Portrait Features	**0.074302**
3	Exposure_Balance	Basic Quality	-0.031212
4	Uniqueness	Fuzzy Properties	**0.14232**
5	Smiling	Portrait Features	-0.045702
6	Cluster4_Lighnting	Compositional	0.017803
7	Fence	Semantics	-0.022525
8	Hue_Inner_Quadrant	Compositional	-0.045009
9	Nose_Hue	Portrait Features	-0.03898
10	Flower	Semantics	0.026438

TABLE II

FEATURE RANKS BASED ON LASSO REGRESSION

Spearman correlation coefficient ρ between the individual features values and the aesthetic scores of each partition.

In Figure 3 (a), we report the ρ coefficients of the features that show higher correlation with portrait aesthetics. We can notice how face sharpness and lighting are of crucial importance for portrait beauty, as suggested by the Lasso analysis of discriminative features, and by portrait aesthetic literature. 4 out of the top 5 positively-correlated features correspond to the landmark sharpness features. Also, the contrast in sharpness between face and background strongly correlates with portrait beauty ($\rho = 0.12$), as well as the *Overall_Sharpness* metric. As hypothesized, lighting patterns are also fundamental for a good portrait. This is shown by the positive ρ of the face/background lighting contrast feature. Moreover, our analysis shows that there is a relation between image beauty and illumination patterns (e.g. Clusters 3 has positive ρ, while Cluster 4 has negative ρ). Overall, our new lighting features show higher relation with beauty than basic brightness features ($\rho = 0.054$ for the *Average_V* features), confirming the need of more complex lighting features for portrait aesthetic evaluation. Similarly, contrast in colors and in gray levels (*GLCM_Contrast* and *Contrast_Michelson*) also show positive correlation with aesthetic scores.

Moreover, negative ρ values for *Noise* and positive correlation with *GLCM_Energy* make us conclude that visually appealing portraits should have a homogeneous, smooth composition without disturbing distortions. We can also see that the amount of *Median_Filtering* is negatively correlated with beauty, showing that too intensive post-processing results in a decrease of the portrait appeal. Surprisingly, *Exposure_Quality* is negatively correlated with beauty, suggesting that playing with over/under exposure results in more appealing pictures. Moreover, negative ρ for some Color Names indicates that beautiful portraits tend to have little regions colored with non-skin colors such as green, purple, magenta. We can also notice the good outcome of our attempt of modeling fuzzy properties, given that properties such as *Originality* and *Uniqueness* positively correlate with beauty.

It was very interesting to notice how physical/demographic properties such as gender, eye color, glasses, age, and race show very low correlation with image beauty, suggesting that any subject, no matter his/her traits, can be part of a stunning picture, if the photographer is able to grasp the subject's essence.

By correlating gender properties with other visual features, we could find some side curious insights about portraiture. For example, female pictures tend to be more memorable, as
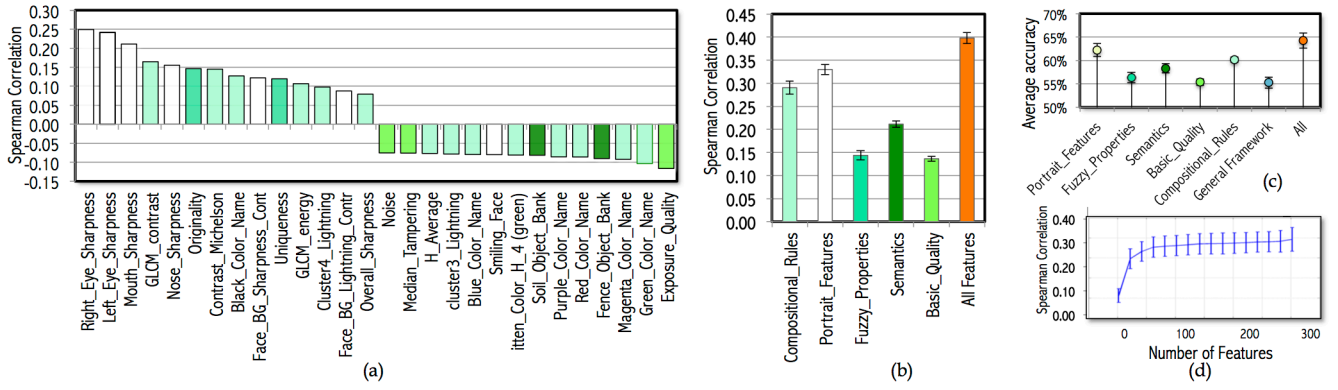
Fig. 3. Analysis of the most relevant features and components for portrait aesthetic prediction (a,b,d). Classification Performances (c).

Feature	Dim	Accuracy
Baseline [7]	7	61.10%
Face Features	44	62,88%
Face Features (sel)	11	68,94%
Non-Face Features	276	65,15%
Non Face Features (sel)	9	68,18%
All Features)	320	66,67%
All Features (sel)	12	75,76%

TABLE III

CLASSIFICATION ACCURACY ON THE DATASET FROM [7]

well as brighter and post-processed, while male tend to be represented with darker colors, and smile less than females.

VI. PREDICTING PORTRAIT BEAUTY

In order to test the effectiveness of our proposed features, and verify the findings of our analysis (see Sec. V), we perform 2 different classification experiments. First, we perform a small-scale experiment on the dataset provided in [7], showing the performances of our method and comparing them with the face-specific framework proposed by [7]. Then, we design a large-scale classification framework, by looking at the ability of our features to discriminate between beautiful/non-beautiful pictures, using the large-scale dataset we built in Sec. III. We compare the classification performances of a framework based on our different groups of features with the one of a generic aesthetic classifier, i.e. based on traditional compositional features and trained on images with diverse subjects.

A. Small-Scale Experiment

The work that more closely relates to ours is the portrait aesthetic framework from Khan et al. [7]: they design face-specific features and computes their effectiveness on a publicly available small-scale dataset of 150 pictures.

In order to test the performances of our approach, we compute the visual features in Sec. IV on the dataset from Khan et al. [7] and we prove their effectiveness by using the same experimental setup, i.e. binarization of scores based on median, 10-folds cross validation on an SVM classifier in Weka, and average accuracy as evaluation metric. For fair comparison, we first evaluate the classification performances on our portrait-specific features only (see Sec. IV-D), reporting results with and without feature selection in Table III. Our group of portrait features alone outperforms the system in [7]. Moreover, when we use all the features proposed

in this work for the same classification task, we reach even higher classification accuracy, observing a substantial improvement of the performances compared to our baseline (and similar works such as the one from Li et al. [8]).

B. Large-Scale Aesthetic Categorization

We now test the proposed approach for aesthetic classification on a large-scale, using the dataset of Sec. III. To classify the images as "Beautiful" and "Non-beautiful", we use the binaries the average AVA scores it by labeling as positive any image with a score greater than the mean user score (5.55). Similar to [9], we learn a SVM classifier using the publicly available libSVM package. For this, the dataset is randomly divided into 5 partitions, as in Sec. V, and a SVM classifier is learned per partitions. We use RBF kernel where the γ parameter is set to $1/n$ where n is the number of features. The cost parameter C is obtained using 10-fold cross-validation. All features are standardized to be zero mean and unit variance.

Fig 3 (c) shows the average classification accuracy on the test set for each group of features. As we can see, our framework benefits from the combination of diverse features, since the best performance is given by all features combined with early fusion, $(64.24\% \pm 1.76)$. Moreover, as expected by our analysis, we confirm that the classifier based on our rich portrait features outperforms the classifiers based on the other groups of features, suggesting that detailed information of face properties and landmarks is more discriminative for portrait classification than traditional compositional features.

Results reported in [9], [14] proved that a classifier trained on non-specific images performs better than a portrait-specific framework. To prove the importance of building a portrait-specific framework, we compare our results with a baseline classifier built with traditional compositional features only (as in Sec. IV-A), and trained on the dataset used in [32], namely a database of images belonging to 7 different categories, including "Portraiture","Flower", etc. and annotated with the corresponding aesthetic scores from DPchallenge.com (same source as our dataset, same score range). Unlike the findings in [9], [14], we confirm the hypothesis that portraits need a separate computational framework for aesthetic assessment, showing that all the classifiers based

on our proposed features perform better than this baseline (with all features, the improvement is more than 16%).

As in [9], we also performed SVM classification by introducing a δ parameter to discard ambiguous images from the training set (keeping all the images in the test set). The δ parameter was ranged from 0.1 to 1.0, but unlike [9] we did not experience any increase in the classification accuracy. However, the performance with the $\delta = 0.5$ is similar to when $\delta = 0.0$, implying that the ambiguous images do not help for the task of classification and can be discarded to speed up the learning time.

VII. Conclusions

In this paper, we presented a complete framework for large-scale portrait aesthetic assessment based on visual features. We procured a dataset of digital portraits annotated with aesthetic scores and other information regarding traits/demographics of the subjects in the portraits. We designed a set of discriminative visual features based on portrait photography literature. We analyzed the importance of each feature for portrait beauty, showing that rich facial features play a significant role in guiding the portrait aesthetics, and that the perceived portrait beauty is largely independent of the demographic characteristics of the subject. Finally, we built a classifier that is able to successfully distinguish between beautiful and non-beautiful portraits.

In our future work, we plan to broaden our framework by extending our database to include portrait images 'in the wild', exploring portrait aesthetics with a more challenging context.

References

[1] Bakhshi, S., Shamma, D., Gilbert, E.: Faces engage us: Photos with faces attract more likes and comments on instagram. (In: CHI 2014)

[2] Child, J.: Studio photography: essential skills. CRC Press (2008)

[3] Hurter, B.: Portrait Photographer's Handbook. Amherst Media, Inc (2007)

[4] Weiser, J.: Phototherapy techniques: Exploring the secrets of personal snapshots and family albums. PhotoTherapy Centre (1999)

[5] Datta, R., Joshi, D., Li, J., Wang, J.: Studying aesthetics in photographic images using a computational approach. ECCV (2006) 288–301

[6] Ke, Y., Tang, X., Jing, F.: The design of high-level features for photo quality assessment. In: CVPR. Volume 1., IEEE (2006) 419–426

[7] Khan, S.S., Vogel, D.: Evaluating visual aesthetics in photographic portraiture. In: CaE, Eurographics Association (2012) 55–62

[8] Li, C., Gallagher, A., Loui, A.C., Chen, T.: Aesthetic quality assessment of consumer photos with faces. In: ICIP, 2010, IEEE (2010) 3221–3224

[9] Murray, N., Marchesotti, L., Perronnin, F.: Ava: A large-scale database for aesthetic visual analysis. In: CVPR, 2012 IEEE Conference on, IEEE (2012) 2408–2415

[10] Isola, P., Xiao, J., Torralba, A., Oliva, A.: What makes an image memorable? In: CVPR, IEEE (2011) 145–152

[11] Machajdik, J., Hanbury, A.: Affective image classification using features inspired by psychology and art theory. In: MM, ACM (2010) 83–92

[12] Joshi, D., Datta, R., Fedorovskaya, E., Luong, Q.T., Wang, J.Z., Li, J., Luo, J.: Aesthetics and emotions in images. Signal Processing Magazine, IEEE 28 (2011) 94–115

[13] Redi, M., Merialdo, B.: Where is the interestingness?: retrieving appealing videoscenes by learning flickr-based graded judgments. In: MM, ACM (2012) 1363–1364

[14] Dhar, S., Ordonez, V., Berg, T.: High level describable attributes for predicting aesthetics and interestingness. In: CVPR, IEEE (2011) 1657–1664

[15] Gygli, M., Grabner, H., Riemenschneider, H., Nater, F., Van Gool, L.: The interestingness of images. (2013)

[16] Zerr, S., Siersdorfer, S., Hare, J., Demidova, E.: Privacy-aware image classification and search. In: ACM SIGIR conference on Research and development in information retrieval, ACM (2012) 35–44

[17] Nishiyama, M., Okabe, T., Sato, I., Sato, Y.: Aesthetic quality classification of photographs based on color harmony. In: CVPR, IEEE (2011) 33–40

[18] Marchesotti, L., Perronnin, F., Larlus, D., Csurka, G.: Assessing the aesthetic quality of photographs using generic image descriptors. In: ICCV, IEEE (2011) 1784–1791

[19] Wu, O., Hu, W., Gao, J.: Learning to predict the perceived visual quality of photos. In: ICCV, IEEE (2011) 225–232

[20] Bhattacharya, S., Sukthankar, R., Shah, M.: A framework for photo-quality assessment and enhancement based on visual aesthetics. In: MM, ACM (2010) 271–280

[21] Luo, W., Wang, X., Tang, X.: Content-based photo quality assessment. In: ICCV), 2011, IEEE (2011) 2206–2213

[22] Xue, S.F., Tang, H., Tretter, D., Lin, Q., Allebach, J.: (Feature design for aesthetic inference on photos with faces)

[23] Inc., M.: Face++ research toolkit. www.faceplusplus.com (2013)

[24] Chakrabarti, A., Zickler, T., Freeman, W.T.: Analyzing spatially-varying blur. In: CVPR, 2010 IEEE Conference on, IEEE (2010) 2512–2519

[25] Michelson, A.: Studies in optics. Dover Pubns (1995)

[26] Rigau, J., Feixas, M., Sbert, M.: Conceptualizing birkhoff's aesthetic measure using shannon entropy and kolmogorov complexity. Computational Aesthetics in Graphics, Visualization, and Imaging (2007)

[27] Li, L.J., Su, H., Xing, E.P., Li, F.F.: Object bank: A high-level image representation for scene classification & semantic feature sparsification. In: NIPS. Volume 2. (2010) 5

[28] Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: CVPR 2005. IEEE Computer Society Conference on. Volume 2., IEEE (2005) 60–65

[29] Wang, Z., Sheikh, H.R., Bovik, A.C.: No-reference perceptual quality assessment of jpeg compressed images. In: Image Processing. 2002. Proceedings. 2002 International Conference on. Volume 1., IEEE (2002) I–477

[30] Borth, D., Ji, R., Chen, T., Breuel, T., Chang, S.F.: Large-scale visual sentiment ontology and detectors using adjective noun pairs. In: ACM international conference on Multimedia, ACM (2013) 223–232

[31] Dan-Glauser, E.S., Scherer, K.R.: The geneva affective picture database (gaped): a new 730-picture database focusing on valence and normative significance. Behavior research methods 43 (2011) 468–477

[32] Obrador, P., Saad, M., Suryanarayan, P., Oliver, N.: Towards category-based aesthetic models of photographs. Advances in Multimedia Modeling (2012) 63–76

[33] Gross, R., Brajovic, V.: An image preprocessing algorithm for illumination invariant face recognition. In: Audio-and Video-Based Biometric Person Authentication, Springer (2003) 10–18

[34] Ng Kuang Chern, N., Neow, P.A., Ang, V.: Practical issues in pixel-based autofocusing for machine vision. In: ICRA. IEEE International Conference on. Volume 3., IEEE (2001) 2791–2796

[35] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR 2005. IEEE Computer Society Conference on. Volume 1., IEEE (2005) 886–893

[36] Hou, X., Zhang, L.: Saliency detection: A spectral residual approach. In: CVPR, IEEE (2007) 1–8

[37] Pevny, T., Bas, P., Fridrich, J.: Steganalysis by subtractive pixel adjacency matrix. information Forensics and Security, IEEE Transactions on 5 (2010) 215–224

[38] : (Credits for the use of the casia image tempering detection evaluation database (caisa tide) v1.0 are given to the national laboratory of pattern recognition, institute of automation, chinese academy of science, corel image database and the photographers. http://forensics.idealtest.org)

[39] Yuan, H.D.: Blind forensics of median filtering in digital images. Information Forensics and Security, IEEE Transactions on 6 (2011) 1335–1345

[40] Redi, M., Merialdo, B.: Saliency moments for image categorization. In: ICMR. (2011)

[41] Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) (1996) 267–288