How to predict the global instantaneous feeling induced by a facial picture?

Arnaud Lienhard, Patricia Ladret, Alice Caplier

How to predict the global instantaneous feeling induced by a facial picture ?

Arnaud Lienhard, Patricia Ladret and Alice Caplier

GIPSA-Lab, Université Grenoble Alpes, France

Abstract

Picture selection is a time-consuming task for humans and a real challenge for machines, which have to retrieve complex and subjective information from image pixels. An automated system that infers human feelings from digital portraits would be of great help for profile picture selection, photo album creation or photo editing. In this work, two models of facial pictures evaluation are defined. The first one predicts the overall aesthetic quality of a facial image, and the second one answers the question "Among a set of facial pictures of a given person, on which picture does the person look like the most friendly?". Aesthetic quality is evaluated by the computation of 15 features that encode low-level statistics in different image regions (face, eyes, mouth). Relevant features are automatically selected by a feature ranking technique, and the outputs of 4 learning algorithms are fused in order to make a robust and accurate prediction of the image quality. Results are compared with recent works and the proposed algorithm obtains the best performance. The same pipeline is considered to evaluate the likability of a facial picture, with the difference that the estimation is based on high-level attributes such as gender, age, smile. Performance of these attributes is compared with previous techniques that mostly rely on facial keypoints positions, and it is shown that it is possible to obtain likability predictions that are close to human perception. Finally, a combination of both models that selects a likable facial image of good quality for a given person is described.

Keywords:
Aesthetic Quality, Likability, Automatic Scoring, Portraits

Email address: `FirstName.LastName@gipsa-lab.grenoble-inp.fr` (Arnaud Lienhard, Patricia Ladret and Alice Caplier)

1. INTRODUCTION

1.1. Context

Social psychological studies have shown that people form impressions from facial appearance very quickly [1]. With the widespread use of digital cameras and photo sharing applications, selecting the best picture of a particular person for a given application is a time-consuming task for humans. Thus, an automated system providing a feedback about facial images would be an interesting and useful tool. Sorting images automatically, editing images to enhance their visual aspect or selecting a few images among an entire collection would be simplified for home users. Generally, images with low aesthetic quality are manually rejected whereas appealing images are selected.

In the particular case of facial pictures, features have to be adapted to the considered use: profile pictures on social networks are different from pictures presented in a professional purpose (resumes, visiting cards). To this end, this work focuses both on predicting the overall aesthetic quality of a facial image and selecting images that infer a feeling of likability. Ideally, the models developed in this work should encode relevant information about the global image aesthetics adapted to facial pictures as well as information related to facial expressions and high-level attributes (smile, age, gender, etc.). Facial beauty is not considered at all: the main idea is to estimate the feeling induced by a given facial picture (especially in term of likability), which is not necessarily correlated with the reality.

1.2. Previous Work

1.2.1. Aesthetic Photo Quality Evaluation

Automated aesthetic evaluation of facial pictures is a challenging task that requires to understand subjective notions that are implicitly encoded in the image. To solve this problem, different approaches exist. In most of recent works, a large number of features describing the image aesthetics are extracted and machine learning algorithms are applied to fit the feature values to ground truth images obtained from human evaluation. Features can either be explored at pixel level (e.g. Fisher Vectors) [2] or by estimation of high-level attributes (smiles, eyes closeness) [3, 4] that are closer to human interpretation. To encode both local and global information into the models,

the main approach for evaluating portraits aesthetic quality is characterized by computing a set of features about the subject (face) and background (non-face) regions. Often, low-level image statistics such as contrast, sharpness or color distribution are computed in addition to features that describe subject-background relationship [5, 6].

To the best of our knowledge, few researches have been done on the particular case of pictures containing a unique and centered frontal face [7]. Plus, there are no publicly available datasets containing facial images and human aesthetic ratings, which makes comparison with previous models difficult. In previous work [8], we developed a method that segments precisely a portrait (hair, shoulders, skin, background) and computes features in each region. The main result of this previous work is that facial area is almost sufficient to describe efficiently the global aesthetic of the entire facial picture. This idea is exploited in the proposed work, where features are extracted in small and informative facial ares (eyes, mouth).

1.2.2. Likability Evaluation

The feeling induced by a facial picture depends on facial expression, face shape and other cues such as make-up or face adornments. However, state-of-the-art face evaluation systems do not consider many of these attributes. A first attempt to create a data-driven model of several evaluation traits is discussed in [9], in which 300 faces are generated by the Facegen Modeller software (http://www.facegen.com) with different shape parameters. A subjective experiment is conducted, where participants evaluate each face with respect to a particular trait: aggressiveness, attractiveness, threat, etc. Finally, shape parameters are fitted to the ground truth scores provided by participants to build a regression model for each social judgment.

Besides, behavioral studies have shown that facial image quality estimation does not only rely in face shape and that reflectance (cues such as skin illumination and texture) also plays an important role in face perception [10]. A more complete model including reflectance parameters is elaborated and validated in [11]. However, the faces considered in all their experiments are synthetic and without facial hair, make-up or accessories. Real 3-D scanned faces have been used in [12] to identify relevant shape and reflectance features. Even in recent attempts of automated face expression evaluation in videos [13], the use of facial keypoints is still predominant. The disadvantage of these models is that it only takes into account the position of facial keypoints and reflectance parameters. Plus, facial keypoints are heavily related

to the face shape whereas our goal is to predict the most likable image of a given person which is not entirely defined by the face shape.

High-level attributes are defined as abstract and global concepts describing an image. They correspond to descriptors that cannot directly be obtained by extracting visual data due to the semantic gap between information contained in pixels and human analysis. Many attributes (age, gender, presence of glasses, beard, smile, etc.) have already been successfully used in various research domains such as face recognition or verification [14] and portraiture aesthetics [3]. A small set of such attributes provides more significant information than the relative positions of many facial keypoints.

1.3. Objectives

The main contribution of this paper is to propose the first model that combines both aesthetic quality assessment and likability estimation for frontal facial pictures, in order to perform automatic picture selection. For each criterion (aesthetic quality and likability), a model that outperforms state-of-the-art methods is presented, and the most relevant features are described.

Aesthetic quality of facial pictures is evaluated using the same feature set than in our recent work [15]. The difference rely on the use of 4 learning algorithms that are combined to provide a more accurate and robust prediction, which outperforms our previous results. This work also focuses on demonstrating the advantages of using high-level attributes in order to build likability evaluation models. 3 tools are considered to compute the attributes: Betaface (*http://betaface.com*), SkyBiometry (*http://skybiometry.com*) and SHORE [16]. It is shown that for real images, these features are significantly more efficient to predict likability.

This work is organized as follows. Section 2 describes the main steps of the proposed method, including feature computation and selection and learning algorithms. Section 3 demonstrates the relevance of the algorithm combination and compares the results with previous recent works. In Section 4, the same pipeline is applied to perform likability evaluation. The major difference between aesthetic quality and likability evaluation is the feature extraction process, which rely either on low-level statistics or on high-level attributes. Finally, both predictions are combined in Section 5 to perform automatic selection that retains automatically good quality images with friendly faces.

2. Proposed Method for Aesthetic Quality Estimation

In this work, headshots are defined as frontal portraits cropped to the extremes of the target's head and shoulders. A method predicting the aesthetic quality of headshots is described in this section. To the best of our knowledge, it reaches the best state-of-the-art results on headshots, and performs well on every dataset containing frontal portraits.

The proposed method consists in computing 15 low-level features in 4 image regions corresponding to the face and its facial attributes (Section 2.1). To eliminate non relevant features, a feature selection algorithm is presented. Then, automated aesthetic prediction is performed by using the learning algorithms that are described in Section 2.2. Prediction performance is evaluated by the metrics defined in 2.2.2 and a fusion technique is proposed in order to combine the advantages of each algorithm in 2.2.3.

2.1. Feature Computation

2.1.1. Facial Attributes Segmentation

Appealing headshots presents a clear architecture: a face located near the image center and well contrasted with respect to the background. Thus, in the proposed model, features of different regions are computed to encode the subject/background relationship. To locate the face area, bounding box detection is performed by using Viola-Jones algorithm [18] and the OpenCV library. Inside the face region, observers are more likely to focus on eyes and mouth, which provide information about the subject: facial expressions, presence of make up, etc. The proposed method relies on the fact that decisive information about facial image aesthetic quality can be obtained by computing features in these small areas only [15] (eyes and mouth). Each image is finally decomposed into the 4 regions described in Figure 1: entire image $\mathcal{R}_A$, face area $\mathcal{R}_B$, eyes area $\mathcal{R}_C$ and mouth area $\mathcal{R}_D$. Eyes and mouth areas are also detected by Viola-Jones algorithm and both eyes are considered to be part of the same region.

2.1.2. Features Extraction

State-of-the-art methods implement a lot of different low-level features (76 in [19], including brightness, contrast, color and sharpness information) in order to assess aesthetic quality of facial images. In this work, the 15 low-level features presented and tested in [15] are considered. They consist in image statistics that can be computed in each region: each face picture is described

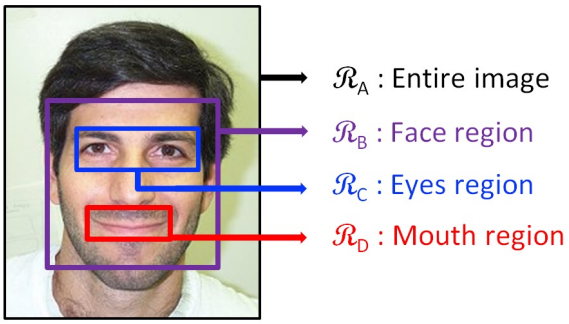

Figure 1: Example of an image and its 4 regions.

by a set of 60 values (15 features in each of the 4 regions). Features correspond to sharpness, illumination, contrast and color distribution measures. These categories have been chosen because they are close to human perception. Thus, it is possible to have feedback about relevant features, which can be helpful for photo editing and feature design. Features are grouped into the 4 categories described below.

Sharpness is evaluated by 3 different values: $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$. The first sharpness measure $\mathcal{F}_1$ is computed by using the blur estimation method described in [20], which compares the difference between an original image I and its low-pass filtered version I_b. More precisely, gradients are measured in I and in I_b: the greater the gradient differences between both images, the sharper the original image I. Indeed, high differences mean that the original image has sharp edges, and loses a lot of its sharpness through the filtering process. On the contrary, blurry images do not change a lot after filtering. This method appeared to be very discriminant in our previous work [8].

Since a sharp facial picture contains high gradients located in the face region, the average gradient value $\mathcal{F}_2$ is computed. The size of the bounding box containing 90% of the image gradients $\mathcal{F}_3$ is calculated as described in [21] in each region.

Illumination is characterized by 2 values, $\mathcal{F}_4$ and $\mathcal{F}_5$, evaluated by the means of two channels: Value V and Luminance L^* (respectively from HSV and $L^*a^*b^*$ color spaces). Both measures are considered in several articles [21, 22, 23]. They provide information about the image global brightness if computed on the entire image, or local brightness if computed on facial regions. Even if these values are highly correlated, both are implemented because the less discriminant measure will automatically be removed by the feature selection process. Combination of local and global measures also gives

some indications about the brightness difference between face and non face regions, which influences our perception of aesthetics [24, 25]. This information is implicitly encoded through the learning process, where brightness values from different regions are fused.

Contrast is measured by 4 values, from $\mathcal{F}_6$ to $\mathcal{F}_9$. Two of them correspond to the standard deviation of V and L^* (respectively $\mathcal{F}_6$ and $\mathcal{F}_7$). Then, the width of the middle 90% mass of L^* histogram $\mathcal{F}_8$ [21, 24] and the Michelson contrast value $\mathcal{F}_9$ [26] are computed. Michelson contrast is obtained by the ratio $(L^*_{max} - L^*_{min})/(L^*_{min} + L^*_{max})$ where L^*_{max} and L^*_{min} are the highest and lowest L^* values in the considered region.

Color information is extracted with the measurement of 6 values, from $\mathcal{F}_{10}$ to $\mathcal{F}_{15}$. The Dark Channel (DC), introduced to perform haze removal [27], provides information about sharpness and colors. High values are related with dull colors or blurry areas. DC corresponds to a minimal filter applied on the RGB color space. Each pixel $p(i, j)$ of an image I is computed as follows: $p(i, j) = \min_{c \in R,G,B}(\min_{(i',j') \in \Omega(i,j)} I_c(i', j'))$ where I_c is a channel of I and $\Omega(i, j)$ corresponds to the 5×5 neighborhood of $p(i, j)$. It has been shown that DC evaluation helps to increase performance of image aesthetic assessment [6]. Since faces are composed of area with low DC values (skin for example) and high DC values (eyes), the DC mean and its standard deviation are considered (respectively $\mathcal{F}_{10}$ and $\mathcal{F}_{11}$).

Hue H and Saturation S standard deviations (from HSV color space) are also computed ($\mathcal{F}_{12}$ to $\mathcal{F}_{13}$). The number of different hues $\mathcal{F}_{14}$ in each area is an indicator of its complexity [21, 3]. Finally, the colorfulness measure $\mathcal{F}_{15}$ described in [28] is implemented, providing information about the mean and standard deviation of the channels a^* and b^* of $L^*a^*b^*$ color space. In recent work [29], it is shown that $\mathcal{F}_{15}$ is highly correlated to the human perception of colorfulness and that this measure is an indicator of the overall image aesthetic quality.

The features described above are represented by numerical values between 0 and 1.

2.1.3. Feature and Region Selection

Some of the considered features may be more relevant when computed in limited regions only. For instance, facial images often have blurred background and sharp edges in the face. Measuring each feature inside all the regions may also add noise in the data due to redundant or irrelevant values. Thus, selecting the most discriminant features for a given area can enhance

the prediction performance, or at least reduce the number of features needed to obtain the optimal result.

In this work, the 60 feature couples (Feature, Region) are ranked using the Relief metric, implemented as described in [30]. This metric provides feedback about the ability of each couple to separate images with similar features but different subjective aesthetic quality scores. The Relief metric is preferred to other feature selection methods because it can be adapted to both classification and regression problems and can compute ranks for each feature couple simultaneously. Plus, it is possible to consider both discrete and continuous features. This will be helpful for likability evaluation in Section 4, for which both types of features are considered.

In case of classification, the Relief value of a given feature f, $Relief(f)_{Class}$, is computed as follows. First, an image i is randomly selected in the training set. For each class, the K closest images to i (measured by the euclidean distance in the feature space) are considered, and the value $Relief(f,i)_{Class}$ is computed:

$$Relief(f,i)_{Class} = \frac{1}{N_c - 1} \sum_{c \neq c_i}^{N_c} \frac{\sum_{k=1}^{K} dF_{ik_c} D_{ik_c}}{\sum_{k=1}^{K} D_{ik_c}} - \frac{\sum_{k=1}^{K} dF_{ik_{c_i}} D_{ik_{c_i}}}{\sum_{k=1}^{K} D_{ik_{c_i}}} \quad (1)$$

where dF_{ik} is the relative difference of feature f for images i and k. D_{ik} is the Euclidean distance between images i and k in the feature space. c_i is the class of image i and N_c corresponds to the total number of classes. K is set to 10, as it has been advised in [30]. The contribution of each nearest neighbor k_c is weighted with respect to the distance between i and k_c.

This process is repeated N_t times, for several images in the training set. Due to the small number of images in the available datasets, in our implementation, the entire training set is considered. For classification, the $Relief$ value is obtained by summing the contribution of each image:

$$Relief(f)_{Class} = \sum_{i=1}^{N_t} Relief(f,i)_{Class} \quad (2)$$

The sum of the first term of Equation 1) represents the probability that two close images (for the Euclidean distance in the feature space) belonging to different classes have close values of f, and the sum of the second term estimates the probability that two close images from the same class have close values of f. As a result, discriminant features are associated to high Relief values, while the value of noisy or irrelevant features are close to 0.

For regression models, it is not possible to consider images that belong to different classes, and the previous formula is adapted as described in [30]. Using the same notations as for Equation 1, the following values are defined:

$$\mathbb{P}_{D_F} = \frac{1}{N_t} \sum_{i=1}^{N_t} \frac{\sum_{k=1}^{K} dF_{ik} D_{ik}}{\sum_{k=1}^{K} D_{ik}} \tag{3}$$

$$\mathbb{P}_{D_S} = \frac{1}{N_t} \sum_{i=1}^{N_t} \frac{\sum_{k=1}^{K} dS_{ik} D_{ik}}{\sum_{k=1}^{K} D_{ik}} \tag{4}$$

$$\mathbb{P}_{D_S,D_F} = \frac{1}{N_t} \sum_{i=1}^{N_t} \frac{\sum_{k=1}^{K} dS_{ik} dF_{ik} D_{ik}}{\sum_{k=1}^{K} D_{ik}} \tag{5}$$

where dS_{ik} is the relative score difference between images i and k. Equations 3, 4 and 5 represent respectively the estimation of the probabilities that two similar images have different values of f (high dF values), different scores (high dS values) and both different scores and different values of f. The expression of $Relief(f)_{Reg}$ is finally:

$$Relief(f)_{Reg} = \frac{\mathbb{P}_{D_S,D_F}}{\mathbb{P}_{D_S}} - \frac{(\mathbb{P}_{D_F} - \mathbb{P}_{D_S,D_F})}{1 - \mathbb{P}_{D_S}} \tag{6}$$

The first term of Equation 6 represents the probability of having similar images with different scores and different values of f, and the second term represents the probability of having similar images with close scores and close values of f. More information about the choice of parameters and implementation details is provided in [30].

2.2. Aesthetic Prediction

2.2.1. Description of the Learning Algorithms

In this work, 4 learning algorithms are considered. They have been chosen due to their ability to perform both classification (separation between low and high aesthetic quality images) and regression (aesthetic quality rating) tasks. These algorithms are implemented in the OpenCV library, and are described below. For each algorithm, OpenCV default parameters are considered, and no significant improvement have been observed by tuning the parameters.

Support Vector Machine. SVM [31] is often used as a classifier in aesthetic prediction [25, 23, 19, 7, 6, 15] and can also be considered for regression analysis [3]. In this work, a Gaussian kernel are used.

Artificial Neural Networks. ANN [32] is a powerful and adaptable algorithm which can be designed for a particular problem by modifying the number of hidden layers and the number of neurons in each layer. In this work, neural networks are built using one hidden layer, containing $N_f/2$ neurons, where N_f is the number of features.

Random Forest. Like SVM, RF [33] is frequently used as a learning algorithm in recent works on automatic aesthetic prediction [25, 19, 34]. It outputs the average of several decision trees (50 in default OpenCV implementation) built using randomly different subsets of the training features and samples. Such a statistical tree-based model makes RF robust to very noisy data, and able to deal with both discrete and continuous features.

Gradient Boosted Trees. GBT [35] is a boosting algorithm using decision trees as weak learners. Thus, it combines the advantages of classical boosting algorithm like Adaboost, and the ability to learn from both discrete and continuous data due to its tree-based structure. OpenCV default parameters are considered: $200 \times c$ trees are built during the learning process, where c is the number of considered classes ($c = 1$ for regression).

2.2.2. Performance Evaluation

In this work, for each experiment, 10-fold cross validation is performed by selecting randomly the testing and training images. This task is repeated 10 times to avoid sampling bias, and only average results are reported.

Classification performance is evaluated by the Cross-Category Error (CCE) and the Multi-Category Error (MCE). Let c_i be the ground truth class and $\hat{c}_i$ the predicted class of image i. CCE is a function of the error magnitude:

$$CCE(k) = \frac{1}{N_t} \sum_{n=1}^{N_t} \chi(c_i - \hat{c}_i = k) \tag{7}$$

where N_t is the number of test images, N_c the number of classes and k is the difference between ground truth and prediction. χ is the function defined by $\chi(x) = 1$ if x is true, $\chi(x) = 0$ otherwise. The Multi-Category Error MCE is the weighted sum of the errors:

$$MCE = \frac{\sum_{k=-(N_c-1)}^{N_c-1} |k| CCE(k)}{MCE_{Rand}} \tag{8}$$

MCE_{Rand} is a normalization constant that is related to the number of classes and corresponds to the MCE value obtained by using a random classifier:

$$MCE_{Rand} = \frac{N_t}{N_c}\frac{N_c^2 - 1}{3} \tag{9}$$

The Good Classification Rate $GCR = CCE(0)/N_t$ is defined as the ratio between the number of images correctly classified $CCE(0)$ and the number of test images N_t. In the case of 2-class categorization, performance can also be measured by the Area Under the ROC Curve AUC. Ideally, GCR and AUC should be the highest (close to 1) and MCE the lowest possible (close to 0).

Regression performance is computed by Pearson's correlation R. Let s_i be the ground truth and $\hat{s}_i$ the predicted score of picture i. R is calculated by the formula:

$$R = \frac{\sum_{n=1}^{N_t}(\hat{s}_i - \bar{\hat{s}}) \cdot (s_i - \bar{s})}{\sqrt{\sum_{n=1}^{N_t}(\hat{s}_i - \bar{\hat{s}})^2} \cdot \sqrt{\sum_{n=1}^{N_t}(s_i - \bar{s})^2}} \tag{10}$$

where $\bar{\hat{s}} = \frac{1}{N_t}\sum_{n=1}^{N_t}\hat{s}_i$ and $\bar{s} = \frac{1}{N_t}\sum_{n=1}^{N_t}s_i$.

2.2.3. Late Score Fusion

For each set of features and data, each learning algorithm will produce different performance. In the case of 2-class categorization, it is possible that one classifier predicts the erroneous class for a particular image, and the 3 other algorithms predicts the correct class. To improve the overall classification performance, the choice has been done to combine the outputs of each algorithm to obtained a fused prediction that corresponds to a weighted vote of each algorithm.

In the case of regression, the combined prediction $\mathcal{P}$ for an image i is defined as:

$$\mathcal{P}(i) = \frac{\sum_{A=1}^{4}\mathcal{P}_A(i) \cdot R_A^p}{\sum_{A=1}^{4}R_A^p} \tag{11}$$

where A is the considered algorithm (SVM, ANN, RF, GBT). R_A is the correlation obtained by cross validation of A on the training set, and $\mathcal{P}_A(i)$ is the quality score predicted by algorithm A for image i.

p is a user-defined parameter that increases or decreases the weight of each algorithm: for $p = 0$, each algorithm will have the same weight and the final prediction corresponds to the mean of each prediction. For a large value of p (e.g. 100), only the algorithm with the highest Pearson's correlation will be considered. In practice, algorithms with low Pearson's values should not be considered, while two algorithms with high values should both be considered. Thus, in all the following experiments, we use $p = 10$ to both remove algorithms that are significantly worse than the best one and retain those producing almost the same performance.

In the case of classification, instead of Pearson's correlation, we consider the GCR and the MCE value to obtain information about the number of images that are correctly classified and the error magnitudes. Since a good classifier presents high GCR and low MCE values, the following formula is used:

$$\mathcal{P}(i) = \frac{\sum_{A=1}^{4} \mathcal{P}_A(i) \cdot (\frac{GCR_A}{MCE_A})^p}{\sum_{A=1}^{4} (\frac{GCR_A}{MCE_A})^p} \tag{12}$$

$\mathcal{P}_A(i)$ is the class predicted by A, and $\mathcal{P}(i)$ is the final prediction. Since classes are designed by integer labels, the overall prediction corresponds to the nearest integer of $\mathcal{P}(i)$. In the following sections, this Late Score Fusion technique is referred as LSF.

3. Validation of the Proposed Method

3.1. Datasets

Experiments are made on 4 different datasets that have been considered for comparison with previous works. Most of state-of-the-art research present results on only one dataset, which can limit the ability of the method to be extended to other type of pictures.

CUHKPQ [6] is a dataset containing 17673 images with manually labeled ground truth, downloaded from the DPChallenge website. Images are separated in seven categories corresponding to different subjects: human, landscape, architecture, etc. For each category, the top and bottom 10% images are labeled respectively as high and low quality images. In this work,

Figure 2: 7 pictures of 3 persons from the HFS dataset.

Figure 3: Samples from the FAVA dataset.

Figure 4: Samples from the Flickr dataset.

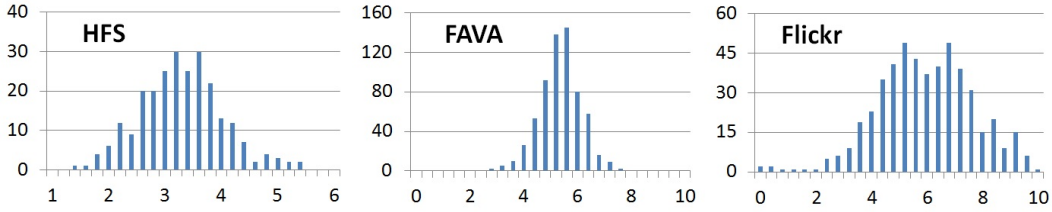

Figure 5: Histograms of ground truth scores for *HFS* (250 images rated from 1 to 6), *FAVA* and *Flickr* (respectively 636 and 500 images rated from 1 to 10).

only the human category is considered, and 335 high quality and 285 low quality headshots are automatically extracted. Due to the significant quality difference between both groups, recent works [2, 6, 34] achieve classification performance above 90%. Since these images are associated to classes and not to scores, it is not possible to perform regression with this dataset.

HFS, for Human Face Scores [8], contains 250 portraits that have been gathered and manually cropped to fit our definition of headshots. More precisely, it contains a set of 7 different images for each of 20 different persons, and 110 additional portraits. Examples of images for 3 particular persons are given in Figure 2. Each image has been rated by 25 persons on a 1 to 6 scale (6 means the highest quality).

FAVA, for Face Aesthetic Visual Analysis, is a subset of the AVA database [36] containing various images from which headshots are extracted. More precisely, each picture is scored from 1 to 10 by internet users (10 means the highest quality). This dataset is similar to the one used in [23] and is used for comparison with previous works. Samples are shown in Figure 3.

Flickr is a website hosting a lot of pictures and portraits. [3] created a dataset of 500 images gathered on this website and scored by the Amazon Mechanical Turk system. Each image is associated to a ground truth score between 0 and 10 (10 means high quality). Photos are either portraits or group portraits. In this work, only the biggest detected face is considered

in each picture, while [3] consider all the faces as well as the relationship between them (distances between faces, face pose and expressions). Images with extremes poses, occluded or very small faces (with a relative size below 1% of the entire image) are automatically removed and finally, 412 pictures are retained for our experiments.

The histograms presented in Figure 5 show that *HFS* presents a higher variance than *FAVA*, for which there is a lot of images with medium scores that makes the learning step more difficult. Thus, prediction performance is likely to be lower for *FAVA* than for *HFS*. Since *Flickr* does not contain only frontal and centered faces but also group portraits, the prediction performance may also be lower than for *HFS*.

3.2. Influence of the Feature Selection Process

15 features and 4 regions ($\mathcal{R}_A$, $\mathcal{R}_B$, $\mathcal{R}_C$, $\mathcal{R}_D$) are a priori considered. Finding the most discriminant couples (Feature, Region) in the case of aesthetic quality estimation presents multiple advantages. First, it helps to design more efficient metrics, adapted to the considered problem. It also enables to compute fewer features, reducing the implementation and computational cost, and finally improving the overall accuracy of the prediction. In this section, the HFS dataset is considered because it contains only centered standardized facial portraits, so that there is no bias related to subject placement or image/face size. For more details about the relevance of each features and regions, refer to [15].

The experimental procedure is the following. Features are computed and sorted using the Relief metric described above. Then, 2-class categorization and regression are performed using SVM on the couple (Feature, Region) with the highest rank. The process is repeated by adding the second most discriminant couple in the model, the third, etc. After the addition of the 60 features, Figure 6 is obtained. 2 observations can be made from this figure. First, less than half of the features are enough to obtain the same performance as the entire set(about 25 for both classification and regression). The second is that by retaining the 35 most relevant features, performance is optimal: GCR and R increase respectively from 86% to 87% and from 0.71 to 0.73. This can be explained by the fact that noisy or less relevant values are removed from the data. In our experiments, the couples with the highest Relief values are $(\mathcal{F}_{\{1,2\}}, \mathcal{R}_{\{B,C,D\}})$: sharpness measures in the facial areas are the most discriminant values for aesthetic quality assessment. This observation is coherent with the model proposed in [29], where sharpness

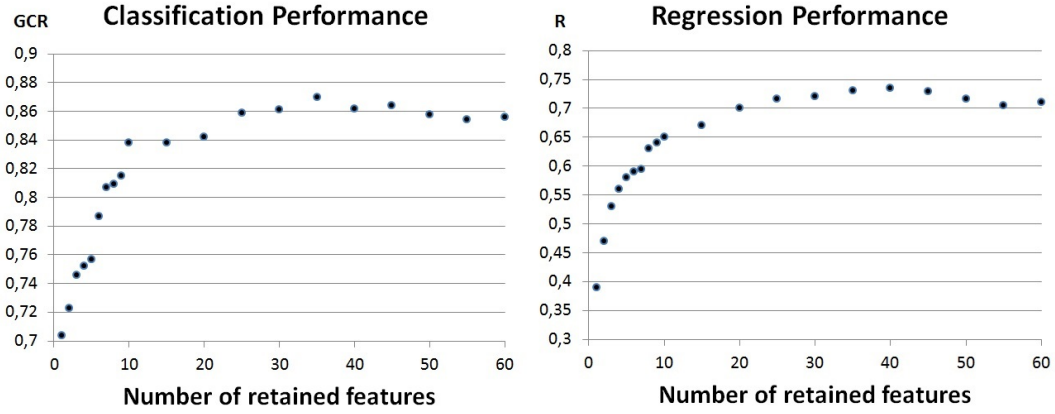

Figure 6: Influence of the number of most discriminant features retained in the model (according to the Relief metric), for 2-class categorization and regression.

metrics are clearly presented as essential because the absence of sharp edges in the subject's area automatically results in low aesthetics ratings.

3.3. Influence of Algorithms

In this section, the entire feature and region sets are considered. 2-class categorization and regression are performed for each of the dataset previously presented. For categorization, datasets are separated in two equally distributed groups (except CUHKPQ which is already separated by labels), containing respectively the images with the lowest and highest aesthetic scores. If SVM generally outputs satisfying performance, in this section it is shown that sometimes it is better to consider different algorithms, and the optimal solution is to combine the algorithms as described in 2.2.3.

For classification, as shown in Table 1, LFS outputs the optimal performance. There are however large disparities among the datasets: CUHKPQ is easily separated ($GCG = 95\%$) whereas FAVA is a difficult dataset ($GCR < 70\%$). This is mostly explained by the fact that there are many average images in FAVA which can be classified in both categories (see Figure 5). Removing 50% of the images with average scores from FAVA leads to approximately 81% of good classification. Pearson correlations R for each dataset are summarized in Table 2. The same observation can be made: LSF enhances performance. CUHKPQ is not considered here since images are not associated to a ground truth score.

Table 1: GCR (%) of the classification algorithms for each dataset.

Datasets	SVM	ANN	RF	GBT	LSF
CUHKPQ	93.9	93.0	90.7	91.4	**94.8**
HFS	77.8	76.6	76.3	76.0	**79.3**
FAVA	65.1	62.7	65.9	66.2	**67.1**
Flickr	69.1	65.2	67.5	67.7	**69.3**

Table 2: Correlation (R) of the regression algorithms for each dataset.

Datasets	SVM	ANN	RF	GBT	LSF
CUHKPQ	/	/	/	/	/
HFS	0.71	0.60	0.69	0.67	**0.73**
FAVA	0.44	0.46	0.47	0.50	**0.51**
Flickr	0.46	0.44	0.43	0.44	**0.49**

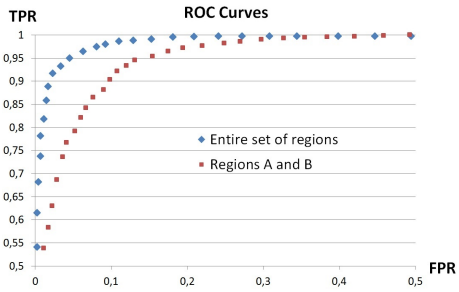

Figure 7: ROC curves obtained by using usual regions (entire image $\mathcal{R}_A$ and face $\mathcal{R}_B$) and the proposed regions, including eyes and mouth.

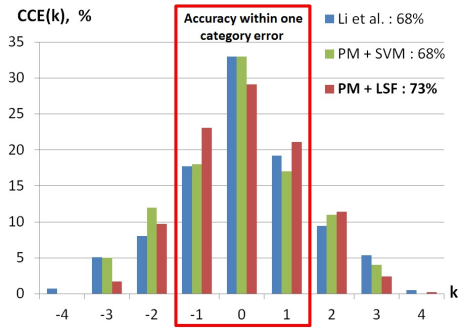

Figure 8: Performance of [3] is compared to the Proposed Method (PM) using either SVM or the LSF.

3.4. Comparison with Previous Works

To compare the proposed method with previous work, the experiments of [3, 23, 25, 8, 34] are reproduced, using the same learning algorithms and databases with the proposed feature set. These works use images containing both group pictures and portraits [3, 34], only portraits [23, 25] or headshots [8, 15]. The method is first compared with previous works performing image categorization, then with works performing score prediction.

3.4.1. Comparison with Previous Categorization Models

CUHKPQ has been considered in several recent works [6, 34] to perform 2-class categorization, and performance is measured by the AUC. [6] and [34] obtain respectively an AUC of 0.974 (SVM) and 0.972 (RF). These results show that this dataset is not very challenging for state-of-the-art methods: only very low and very high quality images are considered, for which very discriminant features exist. Figure 7 presents two curves obtained with the proposed method and SVM classifiers. Without considering eyes and mouth (regions $\mathcal{R}_C$ and $\mathcal{R}_D$ are excluded from the model), performance is very close to [6] and [34] (second curve : $AUC = 0.968$). The first ROC curve is

Table 3: GCR (%) of State-of-the-Art (SotA) is compared with the Proposed Method (PM) for each dataset, using Late Score Fusion.

Dataset	SotA	PM
CUHKPQ	90.0 [6, 34]	**94.8**
HFS	86.5 [15]	**86.9**
Flickr	68.0 [3, 15]	**73.2**
FAVA	**81.0** [15]	**81.0**

Table 4: Correlation (R) of State-of-the-Art (SotA) is compared with the Proposed Method (PM) for each dataset, using Late Score Fusion.

Dataset	SotA	PM
CUHKPQ	/	/
HFS	**0.74** [15]	**0.74**
Flickr	0.47 [15]	**0.49**
FAVA	0.46 [15]	**0.51**

obtained with the entire feature set computed in each region: $AUC = 0.989$. This value is above the results of [6, 34] and it shows, as demonstrated in [15], that computing features in eyes/mouth regions adds significant information to the model in the case of frontal portrait evaluation.

[3] consider 500 images from the Flickr dataset, which are separated in 5 classes with respect to their ground truth aesthetic score. They perform 5-class categorization and measure accuracy within one cross-category error: $(CCE(-1) + CCE(0) + CCE(1))/N_t = 0.68$. Even if our method is not designed to evaluate group pictures, the same accuracy is obtained using the SVM classifier and by performing feature selection. LFS achieves 73% of accuracy as shown in Figure 8, which is above previous performance. This is due to the combination of SVM accuracy ($GCR = 0.32$) and the low error magnitudes obtained from the tree-based classifiers ($MCE < 0.68$ for RF and GBT, 0.71 for SVM). For other datasets, our recent work [15] showed that computing features in additional relevant face regions significantly outperforms methods that are designed to evaluate portraits but do not consider the particular case of headshots. The classification performance for each dataset is compared to state-of-the-art results in Table 3. In some cases, state-of-the-art is not outperformed. This happens when a classifier outputs results that are significantly above other classifiers, so that the combination does not improve the overall performance.

3.4.2. Comparison with Previous Regression Models

Among the 4 works previously cited, only [3] and our previous work [15] performed aesthetic score prediction. [3] calculated the residual sum-

of-squares error RSE to measure performance:

$$RSE = \frac{1}{N_t - 1} \sum_{i=1}^{N_t} (\hat{S}_i - S_i)^2 \tag{13}$$

where N_t is the number of test images, S_i is the ground truth score and $\hat{S}_i$ the predicted score. They perform SVM regression to make score prediction. Using the same dataset and learning algorithm, their features lead to $RSE = 2.38$ while the proposed features lead to $RSE = 2.17$, which is slightly better. RSE can be reduced to 2.12 by combining the algorithms.

Other regression results are directly compared with our previous work [15] in Table 4, showing that most of our previous results are slightly outperformed with the late score fusion.

3.5. Discussion

The results obtained in this section reveal several crucial steps related to aesthetic quality estimation of facial portraits. First, feature extraction and selection are key parts of the process: reduced feature sets can perform equally or better than the entire feature set. By comparing the proposed results with the method developed in [8], it can be observed that it is not necessary to compute precise contours of the face. Simple regions, defined by the face, eyes and mouth bounding boxes add sufficient information to the model. This makes the model easily reproducible with a low computational cost: features can be computed in small regions that are detected in real time via the Viola-Jones algorithm.

If several learning algorithm produce satisfying performance, there is no optimal choice adapted to each feature and dataset. Thus, a trade-off that combines the advantages of the 4 algorithms has been proposed. This step enhances the robustness of the method and is a good alternative to time-consuming parameter tuning that makes the model very dependant. Moreover, it enables to work with both continuous and discrete data, as it is shown in the next section.

Finally, the method described here can be applied to aesthetic quality estimation of other type of pictures, by replacing the face detection step by any other object detection or pyramidal decomposition of a picture. Learning models can also be added in the combination. The proposed method has been limited to the computation of low-level statistics and it is likely that the model would benefit from the addition of other specific features, such as the

lighting templates considered in [25], features that encode directly subject and background relationship [34], or by combining the outputs of each of the feature categories (texture, illumination, contrast, color) as described in [37].

4. LIKABILITY ESTIMATION

The method proposed for facial picture aesthetic estimation is used with a different feature set in order to evaluate how much likable a person looks like on a given picture. Features considered for likability evaluation are high-level descriptors (such as smile, eyes openness, eyebrow position) obtained from the 3 different face analyzers described in Section 4.2. Performance is measured on two types of dataset, involving either synthetic or natural faces. Experiments are made using the feature selection method and the algorithms proposed in Section 2. To the best of our knowledge, few works have explored the automatic evaluation of human traits (trustworthiness, dominance, threat) except for attractiveness due to its application in marketing and cosmetics. In this work, likability estimation is proposed because of its possible application in automatic picture selection for home users, but the same framework can be applied to any other trait.

4.1. Datasets

4.1.1. Synthetic Faces used for Validation

Few datasets containing facial pictures have been built and annotated with respect to likability. In this section, synthetic faces are considered, without any extra-facial cues such as hairstyle, beard, glasses or jewelry.

Using human-rated synthetic facial pictures, models of face evaluation are computed in [9] with respect to the following traits: attractive, competent, dominant, extroverted, likable, threatening, trustworthy. For each model, a dataset of 25 distinct faces is created. These faces are manipulated along the respective traits to generate 7 variations corresponding to 7 different levels of the considered dimension, producing sets of $25 \times 7 = 175$ images. Subjective experiments [11] revealed that these synthetic faces are greatly correlated with the models built in [9]. 7 variations of likability for a given face are presented in Fig. 9. Only the dataset corresponding to likability is considered, and it is used to verify that our evaluation is consistent and to prove that high-level attributes can be good likability descriptors.

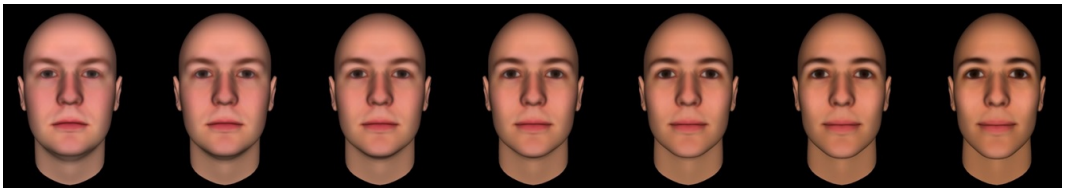

Figure 9: Examples of synthetic faces manipulated for likability [9]. From left to right: unlikable, neutral and very likable.

4.1.2. Human Study on Natural Images

To evaluate our model on natural images, the set of 140 frontal and centered pictures of 20 different persons (10 men and 10 women) from HFS dataset is considered (see Figure 2). To obtain ground truth scores, participants were asked to evaluate each image in the same viewing conditions, and to rate how likable the person on the image seems. Images were presented in a random order, after a preliminary learning process where participants had to rate images that are not part of the dataset. A discrete scale from 1 (not at all likable) to 6 (very likable) has been considered. Finally, 27 participants aged from 20 to 55 rated each image. Scores average and standard deviation are respectively 3.37 and 0.79.

4.2. High-Level Attributes for Face Evaluation

In this work, attributes extraction is performed by 3 tools provided "as is": the SHORE software [16] and two free cloud based applications: Betaface and SkyBiometry. Each tool T returns a total of N_T distinct features. Values may either be discrete (is it a male or a female ?) or continuous (how much is the person smiling ?). Some features have both a discrete component ("yes" or "no") and a continuous component ("how much ?"): Does this person smile (yes or no) ? How much (from 0 to 1) ?

A total of 63 attributes is gathered: 37 from Betaface, 20 from SkyBiometry and 6 from Shore. A simplified list of these attributes is given in Tab. 5. Note that Betaface and SkyBiometry also return a list of facial keypoint positions (respectively 94 and 73 points); detection examples are given in Figure 10. Thus, it is possible to compare keypoints and high-level attributes. Performance of each distinct tool is not discussed in this work (see [17] for more details) and attributes are fused so that each image is described by a vector containing 63 values.

Table 5: Simplified list of high-level attribute categories computed by each tool. The number of values (discrete or continuous) describing each category is reported in each cell.

	Gender	Age	Smile	Mood	Beard	Mustache	Glasses	Eyes	Mouth	Eyebrows	Nose	Skin	Hair	Shape
Betaface	2	1	3		3	3	3	4	3	3	2	2	5	3
SkyBiometry	2		2	8			4	2	2					
SHORE	1	1		1				2	1					

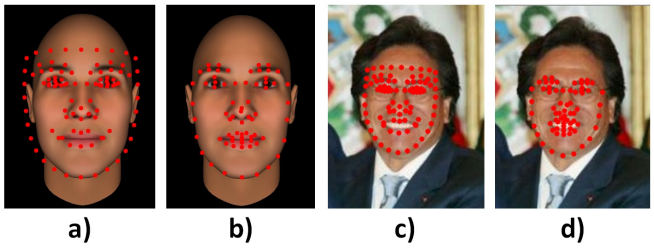

Figure 10: Examples of facial keypoint positions for synthetic and natural images. a) and c) show Betaface points, b) and d) SkyBiometry points.

4.3. Experiments

4.3.1. Validation on Synthetic Faces

In a first set of experiments, synthetic faces are considered to verify that face likability models can be built using the proposed set of high-level attributes. The dataset contains faces grouped in seven categories, corresponding to seven levels of likability (cf Figure 9). Since the faces are generated by keypoints distortion, models based on facial keypoints should provide high classification performance. Our first attempt in creating a likability model requires the facial keypoints provided by either Betaface or SkyBiometry. Applying each learning algorithm, the results presented in the first two lines of Table 6 are obtained.

The performance of a random classifier is approximately $GCR = 14\%$ and $MCE = 1$. Low MCE values reported in the table (between 0.28 and 0.52) and high GCR (between 30 and 50%) indicate that not only our classifier is able to classify correctly many faces (high GCR), but also makes only minor mistakes (low MCE). It is noticeable that SkyBiometry facial keypoints are slightly more efficient than Betaface's: the keypoints detected are not the same for both tools (see Figure 10). Finally, for this configuration of features and images, Neural Networks outperform any other learning algorithms. A possible explanation is that the synthetic faces are very similar, and the high number of keypoints enables the network to fit very well to the learning data. Further experiments on natural images in Section 4.3.2 show that this

Table 6: Experimental results for 7-class categorization using either Betaface / SkyBiometry facial keypoints (KP) or high-level (HL) attributes.

Algorithm	SVM		ANN		RF		GBT		LSF	
Criterium	GCR	MCE	GCR	MCE	GCR	MCE	GCR	MCE	GCR	MCE
KP Betaface	29.9	0.39	33.2	0.40	17.1	0.52	27.7	0.41	35.3	0.38
KP SkyBiometry	34.5	0.33	43.3	0.28	22.5	0.41	33.0	0.35	43.3	0.33
HL Attributes	24.9	0.42	43.8	0.28	22.5	0.41	35.9	0.32	**44.3**	**0.28**

observation cannot be generalized.

The principal contribution of this article in the domain of automatic likability evaluation is the use of high-level attributes. Results are significantly above chance and close to the results obtained with keypoints, even on this particular case of synthetic faces based on keypoint manipulation. This validates the use of HL descriptors as a possible alternative to facial keypoints for likability evaluation. Like the low-level set of features proposed in Section 2, HL attributes provide concrete feedback and can be summarized by a very small number of values that corresponds to human analysis: is the subject smiling ? Is he/she happy ? In the following section, it is shown that in the case of natural images, it is more difficult to evaluate likability by using only facial keypoints.

4.3.2. Application on Natural Images

The dataset of natural images is used for an ultimate validation of the proposed model. Since each picture has a ground truth likability score, it is possible to build regression models. This section validates the use of high-level attributes in the case of natural faces, and shows that facial keypoints are significantly less efficient. In the case of natural pictures, many facial cues play a role in face likability evaluation: hairs, beard, glasses, etc. This is confirmed by the experimental results presented in Table 7. Using attributes instead of keypoints enhances the performance: Pearson's correlation increases from about 20%. In this example, GBT and RF are the most efficient algorithms. ANN is outperformed by tree based algorithms that can efficiently work with features that contain both discrete and continuous data and missing attributes (e.g. eyes that cannot be seen behind sunglasses).

Table 7 shows the correlation between ground truth and predicted scores for the model based on attributes. In the case of natural images, some missing attributes in the feature space (background or reflectance cues are not considered) and errors during the feature extraction process may lead

Table 7: Average correlation (R) for natural images, using either keypoints or combined attributes.

Algorithm	SVM	ANN	RF	GBT	LSF
KP Betaface	0.43	0.43	0.37	0.46	0.50
KP SkyBiometry	0.36	0.27	0.25	0.18	0.39
HL Attributes	**0.74**	**0.72**	**0.81**	**0.82**	**0.84**

to erroneous predictions. This is discussed in the following section, which presents an example of a possible application of this method and its limits.

5. APPLICATION TO AUTOMATED PICTURE SELECTION

Automated picture selection of a given person is a practical application of the proposed method and its results. People may have hundreds of pictures from which they want to select a small set that is relevant for a given application. Since this work focuses on aesthetic quality and likability assessment, possible applications may relate to profile pictures for social networks or meetings websites. The objective is to combine both models (aesthetic and likability) and to sort the pictures with respect to these criteria. The problem is the lack of ground truth data that take both parameters in consideration. Subjective experiments are not a satisfying solution because participants would have to judge simultaneously aesthetic quality and likability. It is possible to ask participants the following question: "Among these images, which one would you choose for this particular application ?". However, this operation ranks the images but does not produce the ground truth scores that are needed for applied the proposed regression models. Thus, in this section, a combination of the aesthetic quality and likability ground truth scores is considered, even if there may be some bias due to the fact that both evaluations are completely different. Several combination are possible (e.g. a weighted linear combination of both ground truth) and the product of both ground is considered in the proposed experiments. Using the product instead of a linear combination ensures that both aesthetic quality and likability evaluations have to be satisfying in order to obtain a high score. The experimental procedure as well as the performance are described in details in Section 5.1. Discussion about the proposed method is given in Section 5.2.

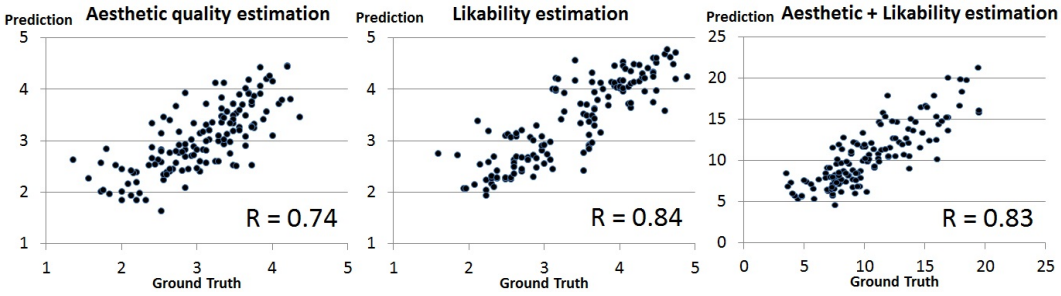

Figure 11: Point clouds obtained from computing with the proposed method, from left to right: aesthetic quality estimation, likability estimation, the product of both evaluations. Ground truth scores of the third image are obtained by multiplying the aesthetic quality and likability ground truth scores.

5.1. Experiments

Since both likability and aesthetic ground truth scores are required, the only dataset that is considered in this section is the subset of HFS containing 7 images of 20 different persons. The objective is to classify pictures for a given person, therefore instead of performing 10-fold cross validation and random sampling, the model is learned using 19 persons and the prediction is made for the remaining 7 pictures of the last person. This task is performed for the 20 different persons, for both aesthetic and likability predictions. Performance is similar to the performance obtained by 10-fold cross validation in the previous sections: the Pearson's correlation for aesthetic quality and likability estimation are respectively 0.75 (0.74 in Section 3.4.2) and 0.85 (0.84 in Section 4.3.2). The small increase of correlation may be due to the use of 20-fold cross validation instead of 10-fold, and in the case of aesthetic quality assessment, only a subset of the entire HFS dataset is considered. Figure 11 presents the point clouds obtained by performing either likability or aesthetic quality estimation and by multiplying the values of both estimations for each picture. The combined prediction, that takes both aesthetic quality and likability estimation into account still presents a correlation of 0.83, implying that it is possible to select pictures that are both of good quality and showing a likable face.

5.2. Discussion

In this section are presented and discussed examples of image automatically selected. Using appropriate thresholds (see Figure 12, it is possible to retain automatically good quality images with likable faces (pictures above

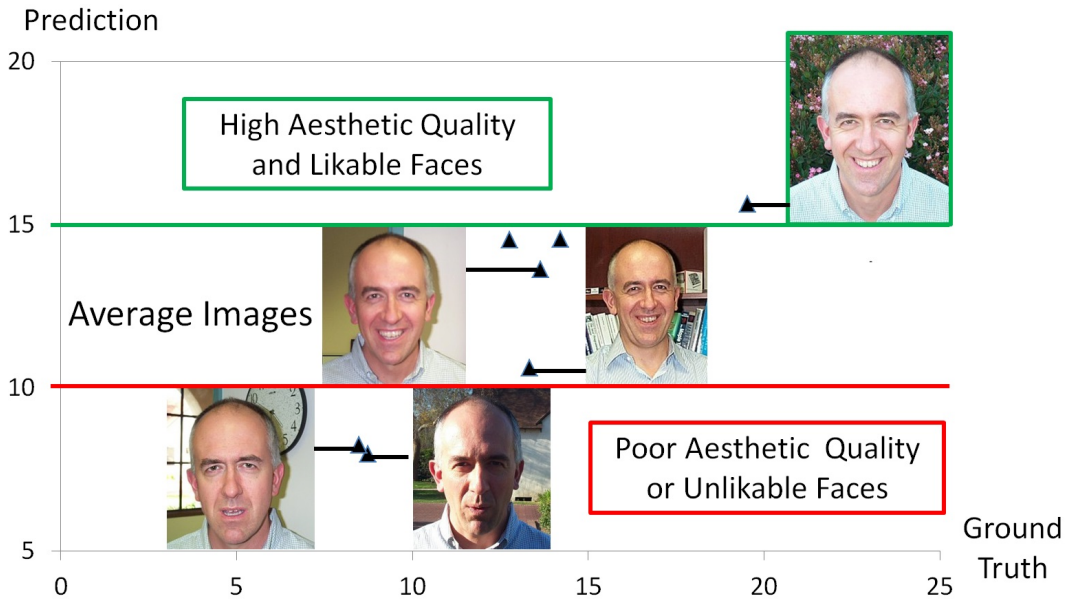

Figure 12: Examples of ground truth and prediction scores for several images of a given person. Images above the green line are selected and images below are rejected.

the green line). Threshold can either be user defined, implying more or fewer accepted images, or by selecting the number of pictures to be selected. This method presents the following drawbacks.

First, it is sensitive to the feature extraction process: it is possible that faces are not correctly detected, or that facial attributes present erroneous measurements. Figure 13 show examples of such mistakes. Face a) is overrated by the aesthetic predictor, which is likely due to the lack of information provided by the color channels. The system has been trained on colored images, as a result the evaluation of an image that does not present the 3 color channels is biased. Faces b) and c) have high likability ratings even if they present non smiling faces: facial expressions are not correctly evaluated for extreme emotions. Rating details are presented in Table 8. Pictures d) to f) are nicely predicted by the model. Note that the system considers closed eyes (picture d)) as satisfying pictures if the person is smiling. This particular case may be solved either by adapting the learning data (which currently does not include poorly rated images with closed eyes) or by performing an additional filtering step where images with closed eyes are removed.

To give more weight to one of the criteria (likability or aesthetics), it is possible to replace the combination function which is currently the product

Figure 13: Images a) to c) are erroneously rated by the system and images d) to f) present correct predictions. Details about prediction values are given in Table 8.

Table 8: Ground Truth (GT) and Prediction (P) results obtained by applying the proposed method on the images presented in Figure 13.

	a)		b)		c)		d)		e)		f)	
	GT	P	GT	P	GT	P	GT	P	GT	P	GT	P
Aesthetics	2.52	2.58	1.72	2.00	1.36	2.63	3.72	3.76	2.64	2.77	3.12	2.58
Likability	1.59	2.75	2.11	3.38	2.63	3.14	3.67	3.62	2.85	2.72	3.26	3.92
Aes. x Lik.	4.01	7.10	3.63	6.79	3.57	8.26	13.6	13.6	7.52	7.55	10.2	10.1

of both predictions. It has to be noticed that the proposed dataset is very limited: only 140 images are considered. The accuracy of the model is both limited by the feature extraction step and the learning data. However, the point cloud presented in Figure 11 show promising applications.

5.3. CONCLUSION

In this paper, a framework for automated facial picture selection has been proposed. To assess aesthetic quality, features are extracted in different face regions (entire face, eyes, mouth) that contain the most relevant information about the portrait. Few pixel-level statistics are computed in each region and late fusion of several learning algorithms is performed to enhance the global prediction performance. Likability is evaluated by extracting different high-level descriptors such as the presence of a smile, eyes closeness or eyebrow position. Then, the process of feature selection and algorithm combination is applied, and satisfying performance is obtained on the considered dataset. A method to combine both predictions is presented, and enables the user to select appealing images, facial images that look likable, or images that fit to both criteria. A major limitation of the proposed likability estimation is that it requires the use of tools that can extract automatically and evaluate correctly high-level attributes.

Results regarding aesthetic quality estimation are promising since recent works are outperformed. The same method is applied to likability evaluation,

and our hypothesis is that the same selection and learning process can be applied to other type of images, by adapting the features and the selected regions. Another possible way to enhance the results would be to add classifiers to the model, and/or to optimize the parameters of the learning algorithms which are currently OpenCV default parameters. It is also possible to employ several times the same algorithm, using different parameters for each time (more trees in the tree based algorithms, different neuron numbers in ANN, etc.). Though, the proposed set of algorithms is quite robust and presents satisfying results.

The choice of the features in this article was motivated by their ability to provide directly feedback to users since they can be easily interpreted: blurry image, poor face enlightening, non smiling face, etc. Plus, the proposed framework can be directly applied for profile picture selection and can be easily applied for other applications, by adapting the learning data and/or the features. In future work, other state-of-the-art features will be added in the model in order to enhance the performance. Also, the feature selection process enables us to decrease the feature number, and by avoiding heavy feature combination it is possible to make the process usable in real-time. Finally, this work focused on likability evaluation, but the framework can be applied to evaluate other traits. For instance, competence evaluation may enable users to select images for professional purposes: resumes, visiting cards, etc. But competence evaluation requires the addition of attributes related to clothes (men in suits are often rated as competent) that are not considered in the proposed model.

[1] J. Willis, A. Todorov, Making Up Your Mind After a 100-Ms Exposure to a Face, Psychological science (2006) 17 (7), pp. 592–598

[2] L. Marchesotti, F. Perronnin, Assessing the Aesthetic Quality of Photographs Using Generic Image Gescriptors, International Conference on Computer Vision (2011), pp. 1784–1791

[3] C. Li, A. Loui, T. Chen, Towards aesthetics: A Photo Quality Assessment and Photo Selection System, International Conference on Multimedia (2010), pp. 827–830

[4] S. Dhar, V. Ordonez, T. Berg, High Level Describable Attributes for Predicting Aesthetics and Interestingness, Computer Vision and Pattern Recognition (2011), pp. 1657–1664

[5] W. Jiang, A. C. Loui, C. D. Cerosaletti, Automatic Aesthetic Value Assessment in Photographic Images, International Conference on Multimedia and Expo (2010), pp. 920–925

[6] X. Tang, W. Luo, X. Wang, Content-Based Photo Quality Assessment, Transactions on Multimedia (2013) 15 (8), pp. 1930–1943

[7] M. Males, A. Hedi, M. Grgic, Aesthetic Quality Assessment of Headshots, International Symposium ELMAR (2013) , pp. 89–92

[8] A. Lienhard, M. Reinhard, A. Caplier, P. Ladret, Photo Rating of Facial Pictures based on Image Segmentation, International Conference on Computer Vision Theory and Applications (2014), pp. 329–336

[9] A. Todorov, N. Oosterhof, Modeling Social Perception of Faces, Signal Processing Magazine (2011) 28 (2), pp. 117–122

[10] V. Blanz, T. Vetter, A Morphable Model for the Synthesis of 3D Faces, Annual Conference on Computer Graphics and Interactive Techniques (1999), pp. 187–194

[11] A. Todorov, R. Dotsch, J. M. Porter, N. N. Oosterhof, V. B. Falvello, Validation of Data-driven Computational Models of Social Perception of Faces, Emotion (2013) 13 (4), pp. 724–738

[12] M. Walker, T. Vetter, Portraits made to measure: Manipulating Social Judgments about Individuals with a Statistical Face Model, Journal of Vision (2009) 9 (12), pp. 1–13

[13] D. Masip, M. S. North, A. Todorov, D. N. Osherson, Automated Prediction of Preferences Using Facial Expressions, PloS one (2014) 9 (2), pp. e87434

[14] N. Kumar, A. C. Berg, P. N. Belhumeur, S. K. Nayar, Attribute and Simile Classifiers for Face Verification, International Conference on Computer Vision (2009), pp. 365–372

[15] A. Lienhard, P. Ladret, A. Caplier, Low Level Features for Quality Assessment of Facial Images, International Conference on Computer Vision Theory and Applications (2015)

[16] A. Ernst, T. Ruf, C. Kueblbeck, A Modular Framework to Detect and Analyze Faces for Audience Measurement Systems, Workshop on Pervasive Advertising at Informatik (2009), pp. 75–87

[17] A. Lienhard, P. Ladret, A. Caplier, Fully Automated Facial Picture Evaluation Using High Level Features, Automatic Face and Gesture Recognition (2015)

[18] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, Computer Vision and Pattern Recognition (2001) 1, pp. 511–518

[19] J. Faria, S. Bagley, S. Rüger, T. Breckon, Challenges of Finding Aesthetically Pleasing Images, International Workshop on Image Analysis for Multimedia Interactive Services (2013) 2, pp. 4–7

[20] F. Crete, T. Dolmiere, P. Ladret, M. Nicolas, The Blur Effect: Perception and Estimation with a New No-reference Perceptual Blur Metric, Human Vision and Electronic Imaging (2007) XII (6492)

[21] Y. Ke, X. Tang, F. Jing, The Design of High-level Features for Photo Quality Assessment, Computer Vision and Pattern Recognition (2006), 1, pp. 419–426

[22] R. Datta, D. Joshi, J. Li, J. Wang, Studying Aesthetics in Photographic Images using a Computational Approach, European Conference on Computer (2006), pp. 288–301

[23] D. Pogačnik, R. Ravnik, N. Bovcon, F. Solina, Evaluating Photo Aesthetics using Machine Learning, Data Mining and Data Warehouses (2012), pp. 197–200

[24] L.-k. Wong, K.-l. Low, Saliency-enhanced Image Aesthetics Class Prediction, International Conference on Image Processing (2009), pp. 997–1000

[25] S. Khan, D. Vogel, Evaluating Visual Aesthetics in Photographic Portraiture, Annual Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging (2012), pp. 55–62

[26] M. Desnoyer, D. Wettergreen, Aesthetic Image Classification for Autonomous Agents, International Conference on Pattern Recognition (2010), pp. 3452–3455

[27] K. He, J. Sun, X. Tang, Single Image Haze Removal Using Dark Channel Prior, Transactions on Pattern Analysis and Machine Intelligence (2011), 33(12), pp. 2341-2353

[28] D. Hasler, S. Suesstrunk, Measuring Colorfulness in Natural Images, Electronic Imaging (2003), pp. 87–95.

[29] T. Aydin, A. Smolic, M. Gross, Automated Aesthetic Analysis of Photographic Images, Transactions on Visualization and Computer Graphics (2013), 21 (1), pp. 31–42

[30] M. Robnik-Šikonja, I. Kononenko, Theoretical and empirical analysis of ReliefF and RReliefF, Machine learning (2003), 53 (1-2), pp. 23–69

[31] C. Chang, C. Lin, LIBSVM: A Library for Support Vector Machines, Transactions on Intelligent Systems and Technology (2011), 2 (3), pp. 1–27

[32] Y. LeCun, L. Bottou, G. Orr, K. Müller, Efficient Backprop, Neural networks: Tricks of the trade (2012), pp. 9–48

[33] L. Breiman, Random Forests, Machine learning (2001), 45 (1), pp. 5–32

[34] J. Kim, C. Kim, Aesthetic Quality Classification via Subject Region Extraction, International Conference on Image Processing (2014), pp. 536–540

[35] J. Friedman,Stochastic gradient boosting, Computational Statistics & Data Analysis (2002), 38 (4) (2002), pp. 367–378

[36] N. Murray, L. Marchesotti, F. Perronnin, AVA: A Large-scale Database for Aesthetic Visual Analysis, Computer Vision and Pattern Recognition (2012), pp. 2408–2415

[37] L. Zhang, Y. Gao, C. Hong, Y. Feng, J. Zhu, D. Cai, Feature Correlation Hypergraph: Exploiting High-order Potentials for Multimodal Recognition, Transactions on Cybernetics (2014), 44 (8), pp. 1408–1419.