Assessment of female facial beauty based on anthropometric, non-permanent and acquisition characteristics

Article *in* Multimedia Tools and Applications · September 2014

DOI: 10.1007/s11042-014-2234-5

Antitza Dantcheva
INRIA Méditerranée

Jean-Luc Dugelay
EURECOM

Multimed Tools Appl (2015) 74:11331–11355
DOI 10.1007/s11042-014-2234-5

Assessment of female facial beauty based on anthropometric, non-permanent and acquisition characteristics

Antitza Dantcheva · Jean-Luc Dugelay

Received: 9 May 2013 / Revised: 3 July 2014 / Accepted: 11 August 2014 /
Published online: 9 September 2014

Abstract In this work we study the interrelation between, on the one hand, subjective perception of female facial aesthetics, and on the other hand, selected objective parameters that include facial features, photo-quality, as well as non-permanent facial characteristics. This study seeks to provide insight on the role of this specific set of features in affecting the way humans perceive facial images. The approach is novel in that it jointly considers both previous results on photo quality and beauty assessment, as well as non-permanent facial characteristics and expressions. Based on 37 such objective parameters, we construct a metric that aims to quantify modifiable parameters for aesthetics enhancement, as well as tunes systems that would seek to predict the way humans perceive facial aesthetics. The proposed metric is evaluated on a face dataset, that includes images with variations in illumination, image quality, as well as age, ethnicity and expression. We show that our approach outperforms two state of the art beauty estimation metrics. In addition we apply the designed metric in three interesting datasets, where we assess beauty in images of females before and after plastic surgery, of females across time, as well as of females famous for their beauty. We conclude by giving insight towards beauty prediction.

Keywords Facial beauty · Facial aesthetics · Facial symmetry · Female aesthetics · Female beauty · Beauty prediction · Beauty estimation · Beauty assessment · Plastic surgery · Aging beauty

1 Introduction

With millions of images appearing daily on Facebook, Picasa, Flickr, or on different social and dating sites, photographs are often seen as the carrier of the first and deciding

A. Dantcheva (✉)
INRIA, Sophia Antipolis, France
e-mail: antitza.dantcheva@inria.fr

J.-L. Dugelay
EURECOM, Route des Chappes, France
e-mail: jean-luc.dugelay@eurecom.fr

impression of a person. At the same time though, human perception of facial aesthetics in such photographs is a priori highly subjective. The nature of this perception has been explored *separately* from photographical as well as psychological perspective, focusing either on the properties of the image, or on the other hand on the characteristics of the subject. The photographical point of view, often interested in photo-quality and -aesthetics assessment and also enhancement, has recently attracted further attention [3, 12, 61, 62], partly due to the vast amount of digital images that are now available, as well as due to the ease with which digital image modification can now be achieved.

1.1 Related work

The present work draws from former work in three areas, namely classical facial aesthetics, photo–quality and aesthetics as well as image processing based face recognition.

1.1.1 Facial aesthetics

There are substantial amounts of works, both from psychological and sociological perspective, studying human perception of facial attractiveness and beauty [58]. Such perception is highly subjective and is influenced by sociological and cultural factors and furthermore by individual preferences. In attempts to rationalize subjective perception of facial beauty, many objective characteristics have been associated, such as the *golden ratio* [23], facial *symmetry* [44], *neonate features* [69], *averageness* [40] [50], facial *skin texture* [27], hair color, femininity, *familiarity* [38, 60], as well as geometrical attributes of facial features [69]. We proceed to elaborate on the most common characteristics.

– The *golden ratio*: $\Phi \approx 1.618$. When this divine proportion appears in both nature or art, they are perceived harmonic and aesthetic, see [23]. An attractive human face contains Φ in several proportions [51], e.g face height / width and face height / location of eyes, see Fig. 1. Such proportions are pertinent across ethnicities [45].
– *Symmetry* was evolutionary beneficial for its direct analogy to health [44]. We note here that there are contradictory hypotheses regarding the role of symmetry in beauty perception [31].
– *Averageness* is an indicator for health and fertility. In [40] the authors present a study showing that mathematically averaged faces are considered beautiful. Interestingly, exceptionally beautiful faces though are not close to average [2]. This aspect is carried over in other studies, that claim that attractiveness implies distinctive facial features. Such features are named to be: a narrow face, full lips, long and dark eyelashes, high cheek bones and a small nose [65]. An overview of psychological studies based on face proportions and symmetries is presented in [4].
– Finally the neonate features or babyfaceness elicit sympathy and protective urges. Traits in babyfaces include a big round forehead, low located eyes and mouth, big round eyes and small chin, see Fig. 2. In a related study [36] juvenilised faces were examined to be perceived highly attractive.

Despite the continuous interest and extensive research in cognitive, evolutionary and social sciences, modeling and analysis of human beauty and aesthetic canons remains an open problem.

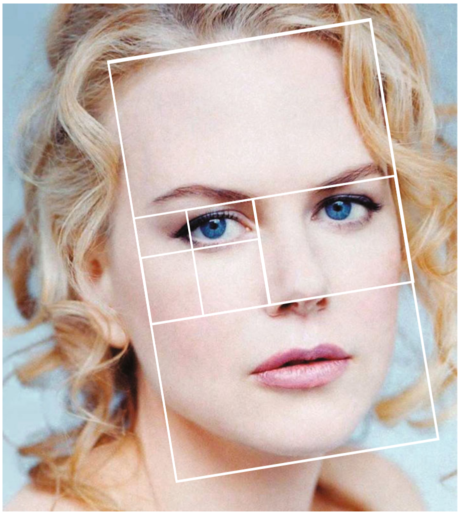

Fig. 1 Golden ratio appears multiple times in a beautiful human face. The ratios (face height / face width) and (face height / location of eyes) are both equal to $\Phi \approx 1.618$. Image obtained from http://nicolekidmanofficial.com/

1.1.2 Beauty and image processing

Recently, pattern analysis has allowed for novel perspectives and insightful analysis regarding human perception of facial attractiveness [4]. Often such analysis aims to map a ground truth for beauty perception - such as a mean value of recorded human ratings - onto extracted facial features. This is often combined with a subsequent supervised learning of underlying interrelation patterns. Such analytical studies consider a variety of extracted facial features and classification algorithms, as well as different datasets, different numbers of human raters, different measures of beauty, as well as a plethora of selected features and used classifiers (see Table 1). A good review of recent works can be found in [31, 41].

The most common facial features considered for beauty assessment are related to distances between geometrical facial landmarks [21, 24]. For example, Mao et al. [43]

Fig. 2 Babyfacesness theorem: females with child like traits including big and low located eyes, big round forehead and small chin, are perceived attractive

Table 1 Review of scientific work on beauty assessment

Work	Dataset	Nr. of images	Cat.	Performance	Nr. of Raters
Gunes et al. [32]	private	215F	5	error rate: stand. dist. sd = 0.68 − 0.85	48
Kagian et al. [37]	from [24]	184F, Caucasian	7	$r = 0.82$, $p < 10^{-23}$	28
Aarabi et al. [1]	private	80F	4	class. rate 91 %	12
Mao et al. [43]	CAS-PEAL [29], WWW	510F, Chinese	2/4	class. rate 95.3 %/77.9 %	N.A.
Türkmen et al. [34]	private	150F	2	class. rate 89 %/83 %	50
Chang and Chou [8]	private	90F, 70M	7	class. rate 97.1 %, 100 %	31M, 28F
Davis and Lazebnik [21]	HOTorNOT	2097F, 1523M	10	web images[a]	HOTorNOT
Said and Todorov [53]	Facegen[b]	2100M, 2100F	9	$r = .84$, $p < .05$, $r = .79$, $p < .05$	20F, 20M
Eisenthal et al. [24]	private	184F, Caucasian	7	$r = 0.45/0.6$	28/18
Gray et al. [30]	HOTorNOT	2056F	10	$r = 0.458$	30

Used abbreviations: Cat...categories, F...female, M...male, r...Pearson's correlation coefficient. If the performance is reported in percentage, it refers to a correct classification rate

[a] http://www.cs.unc.edu/?davisb/research/Regression-Attractiveness/FaceRegression.html

[b] http://www.facegen.com

extracted 17 such geometric features, which they used - in conjunction with support vector machines (SVMs) - to classify 2 beauty categories (attractive, non-attractive), achieving classification accuracy of 95.3 %, and then to classify 4 beauty categories (very attractive, very non-attractive and 2 intermediate categories), with accuracy of 77.9 %. Similarly, Gunes et al. [32] focused on geometric ratios that are classified by decision trees, a Multi-Layer Perceptron and Kernel Density Estimators. Aarabi et al. [1] extracted 8 geometric ratios between facial landmarks (eyes, brows and mouth) and used k-nearest neighbour (k-NN) for classification of 4 categories, achieving 91 % correct classification on a 80 images dataset. Other landmark-based approaches have been discussed in [10, 13, 24, 34, 69].

The dataset from [24] was used by Kagian et al. [37], where 84 facial landmarks were extracted (describing the locations of eyes, nose, lips, eyebrows and head contour) and the associated distances and slopes of the lines connecting facial landmarks were computed. In addition hair and skin color (HSV values) and skin texture (smoothness) were extracted. Principal component analysis was instrumental in reducing of dimensionality, and linear regression was used to build a metric, which achieved a prediction accuracy of 0.82 of the Pearson's correlation coefficient.

Deviating from the above approach, Chang and Chou [8] assessed beauty using a relative distance from two prototypes, namely, the average of attractive faces and the average of unattractive faces. For describing faces, 46 features were annotated which were averaged over attractive and over unattractive faces. An accuracy of up to 100 % was reached, using a relatively small and constrained dataset of 90 females and 70 males. Gray et al. [30] presented a fully automated and holistic beauty prediction algorithm based on multi-layer neural network. The used dataset included 2056 female images, and it incorporated variations in pose, lighting, background, expression, age and ethnicity. The multi-layer feature extractor sought to emulate the way a human brain would holistically extract the features. The authors also showed the pertinence of a new beauty trait; a high contrast in the eye region. Said and Todorov [53] performed experiments on a synthetic dataset of 2100 females and 2100 males. The face generator initially introduced variations in shape and reflectance on the generated faces, and the corresponding model was based on 25 shape dimensions and 25 reflectance dimensions. The first shape dimension was face width, the first reflectance dimension was darkness / lightness. Second-order polynomial regression then mapped these features onto an 'attractiveness' value. The experiments achieved a Pearson's correlation coefficient of 0.79.

Different studies went beyond predicting a mean opinion score for beauty. For example, Whitehill and Movellan [66] modeled individual preferences of 8 subjects, using a model based on PCA, Gabor wavelets and Gaussian RBFs, as well as expression extraction. The model reached a Pearson's correlation rate of up to 0.45. Davis and Lazebnik [21] built a face-shape-model and used manifold kernel regression to 'deform' over 3000 faces towards attractiveness. Chen and Zhang [10] presented a benchmark dataset for female and male faces containing 15,393 Chinese females and 8,019 males. In addition they extracted geometric features and projected them onto a tangent space (a linear space where Euclidean distance measures different facial shapes). They used PCA and revealed that first principal components corresponded to variation in the face width, whereas second principal components corresponded to variations of eyebrow length, and third principal components corresponded to the location of facial features. This study suggested that neither symmetry nor averageness played prominent roles.

Suitability of quality measures in beauty assessment The most common quality measure in beauty prediction is the Pearson's correlation coefficient (PCC), see Section 3.1. PCC is a simple and low-complexity measure, quantifying the correlation between two variables, associated to several conditions of use. Such conditions include a normal distribution of the two variables, linear correlation between the two variables and homoscedasticity (similar standard deviations of both variables) [39]. It is of interest to note that PCC is sensitive to outliers and thus such need to be extracted.

For our main result, we report PCC, Spearman's Rank Coefficient and MSE to thoroughly examine linear relationship, monotonic (and eventually non-linear) relationship and standard deviation of the values between MOS and $\widehat{MOS}$. In Fig. 3 we display the MOS-histogram of a large dataset containing 2056 images, assembled from HOTorNOT.com, as presented in [30]. We observe that the histogram forms a normal distribution. In Fig. 4 we illustrate the MOS-$\widehat{MOS}$ relationship related to our test set. Here, we observe a nearly linear relationship between the two variables.

1.1.3 Photo–quality and aesthetics

Broad background work on image quality assessment (IQA) is needed in applications such as image transmission, lossy compression, restoration and enhancement. The subjective criteria intertwined with image quality are assessed in numerous metrics for mobile phones, electronic tabs or cameras. A number of automatic IQA algorithms has been built based on those metrics. For an overview of related works, see [63] and [55].

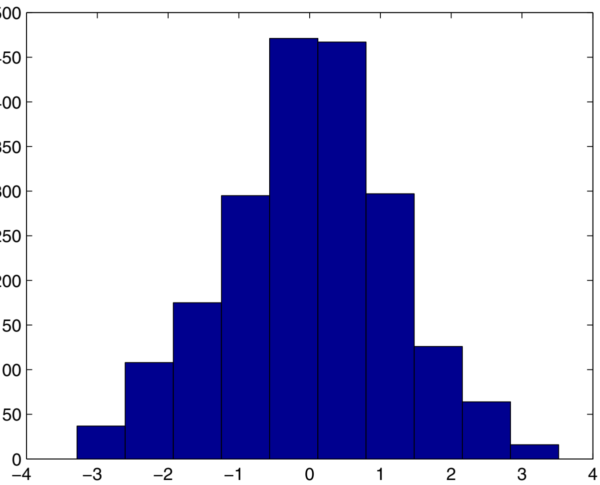

Fig. 3 MOS-histogram of a large dataset containing 2056 images, assembled from HOTorNOT.com for beauty assessment [30]. The related MOS has been normalized and is thus not of a range $1 - 10$

From a photographic point of view, the presence of people, their facial expressions, image sharpness, contrast, colorfulness and composition are factors which play a pivotal role in subjective perception and accordingly evaluation of an image, cf. [54]. Recent works on photography considerations include [3] and [12]. Hereby the authors reveal that appealing photographs draw from single appealing image regions and their location and use this proposition to automatically enhance photo-quality. Photo-quality can also be influenced by image composition, see [49]. Finally there are current studies, which model aesthetic perception of videos [46]. Such methods have become increasingly relevant due to the prevalence of low price consumer electronics.

1.2 Contribution

In this work we study the role of objective measures in modeling the way humans perceive facial images. In establishing the results, we incorporate a new broad spectrum of known aesthetical facial characteristics, as well as consider the role of basic image properties and

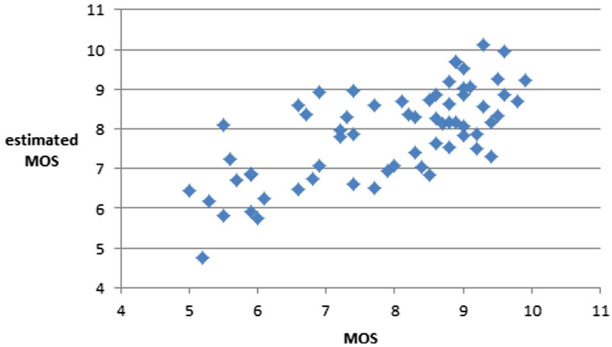

Fig. 4 MOS-$\widehat{MOS}$ relationship related to our test set

photograph aesthetics. This allows us to draw different conclusions on the intertwined roles of facial features in defining the aesthetics in female head-and-shoulder-images, as well as allows for further insight on how aesthetics can be influenced by careful modifications.

Towards further quantifying such insights, we construct a metric that models the role of selected traits in affecting the way humans perceive such images. This model applies as a step towards an automatic and holistic prediction of facial aesthetics in images.

The proposed metric accepts applications, answering following questions:

- How to photograph a face, maximizing the perceived facial attractiveness (e.g. by enhancing illumination, selecting facial expression and makeup)?
- How to enhance or retouch an image, maximizing facial attractiveness?
- Which facial image to choose for an online presentation (e.g. CVs, Facebook, homepage, dating sites)?
- What is an expected beauty improvement after a plastic surgery?

In this work we use the proposed metric to examine facial aesthetics in the context of aging, facial surgery and a comparison of average females relatively to selected females known for their beauty.

The novelty in our work lies mainly in two aspects. The first one is that we expand the pool of established facial features to include non-permanent features such as makeup, presence of glasses, or hair-style. The second novelty comes from the fact that we seek to combine the results of both research areas, facial aesthetics and photograph aesthetics, and thus to jointly study and understand the role of facial features and of image processing states.

The current work is an extension of the conference paper [14]. Here we extend the experiments by

- providing results on 3 additional datasets (aging, facial surgery and famous females),
- comparing the presented metric with two reference state-of-the-art algorithms (one feature-based and one holistic) on the presented dataset,
- including a comparison of different regression techniques,
- presenting additional results on the presented dataset (crossvalidation),
- positioning our work within the field of beauty estimation.

The paper is organized as follows. In Section 1.1 we give a broad overview of facial aesthetics and associated image processing-related work. In Section 2 we introduce the assembled and employed dataset, as well as describe the basic features and methods used in this study. In Section 3 we proceed with numerical results, and provide intuition on the role of features, image quality and facial features, in human perception. In Section 4, we use these accumulated conclusions to construct a model that predicts attractiveness in facial photographs using different facial traits as well as image properties. We then examine, validate and compare the designed metric with two state of the art beauty assessment techniques. In Section 5 we employ the developed metric to conduct experiments and answer questions regarding beauty index in three cases: for famous attractive females, for aging females and in case of facial surgery.

2 Study of aesthetics in facial photographs

In our study we consider 37 different characteristics ($x_1 - x_{37}$) that include facial proportions and traits, facial expressions, as well as image properties. All these characteristics are, for the most part, manually extracted from a dataset of 325 facial images. Each image is associated with human ratings for attractiveness, as explained in Section 2.1. The dataset forms the empirical base for the further study on how different features and properties relate to attractiveness. We proceed with the details of the dataset and characteristics.

2.1 Dataset

We have assembled a dataset including 325 random head-and-shoulders images from the web site HOTorNOT [20]. HOTorNOT has been previously used in image processing studies (cf. [30, 57]), due to the sufficiently large library of images, and the related beauty ratings and demographic information.

Each image depicts a young female subject (see for example Figs. 5 and 6.) and was rated by a multitude of users of the web site. The rating, on a scale of one to ten, corresponds to the notion of attractiveness. The relevance and robustness of the provided ratings was confirmed in an experiment in [42], where subjects re-rated a collection of HOTorNOT-images. For increasing robustness, we consider only images that have received a sufficiently high number of ratings, specifically more than 70 ratings. We will henceforth refer to the mean value of these ratings as the *Mean Opinion Score* (MOS). Among the chosen images of the dataset the mean MOS was 7.9, the standard deviation was 1.4, whereas the minimum value was 4.3 and the maximum value was 9.9. The JPEG images in the here presented dataset are of different resolution and of different quality.

We now proceed with the description of the two sets of considered features: the photograph-aesthetics (image properties) and the facial aesthetics. All characteristics from these sets are stated in Table 2.

2.2 Photograph aesthetics

The considered photograph aesthetic features are here specifically chosen to be simple and objective. Firstly we include characteristics such as image resolution (x_{20}), image format (x_{14}, with possible values describing either portrait or landscape) and illumination (x_{27}). Furthermore, we consider the relative angle of the face in the photograph (x_{34}), being the angle between the line connecting the eyes and the horizontal line (this angle is denoted as α and is illustrated in Fig. 5). We also incorporate the zoom-factor (x_{31}), specifically how large the face appears in comparison to the image height. Finally we also connect three novel image quality traits with facial aesthetics, which in previous work have been associated to photograph-aesthetics: the *relative foreground position*, the *BIQI* (blind image quality index) and the *JPEG quality measure*.

We define the *relative foreground position* to be, as in [3], the normalized Euclidean distance between the foreground's center of mass, also called visual attention center, to each of the symmetric stress points (strongest focal points in a photograph, see [35], indicated by the four white points in Fig. 5). Specifically for our study, we compute and compare the distances of the foreground's center of mass (left eye for x_{17}, right eye for x_{18} or nose tip for x_{26})[1] to any of the stress points (cf. [32]) with the distances of the same foreground's center of mass to the center of the image. For clarity Fig. 5 illustrates the stress points of the image, where each of the four stress points is in a distance of one third the image width and one third the image height from the boundary of the image, an aspect derived from the "Rule of thirds". In case that the foreground's center of mass is equidistant to all stress points,

[1]Culture Shapes www.wired.com/wiredscience/2008/08/culture-shapes/

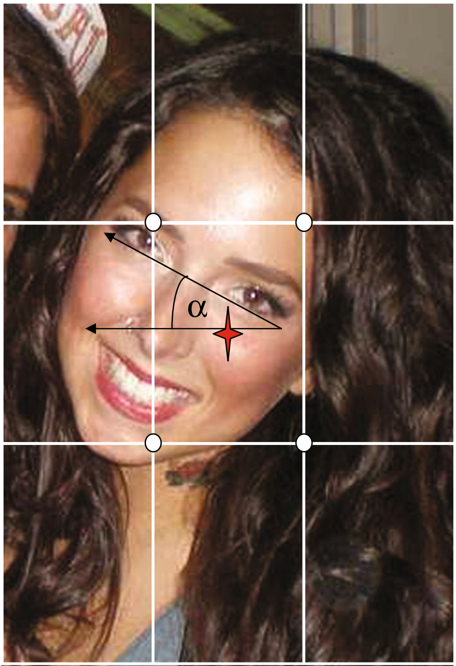

Fig. 5 Example image of the web site HOTorNOT, $MOS = 9.8$. The white points represent the stress points, the *red* cross the image center. The *relative foreground position*, as in [3], is the normalized Euclidean distance between the foreground's center of mass, to each of the stress points. Specifically for our study, we compute and compare the distance of the foreground's center of mass (*left* eye for x_{17}, *right* eye for x_{18} or nose tip for x_{26}) to any of the stress points with the distance of the same foreground's center of mass to the center of the image. The relative angle of the face α comprises the angle between the line connecting the eyes and the *horizontal line*

which is the case in the image center, it has been shown that subjects lose their attention and interest.

The *BIQI measure* (x_{35}) is an image quality measure, based on the distorted image statistics and it employs support vector machines for classification (cf. [47] and [48]). It is a blind quality measure; specifically it is a no-reference assessment measure on image quality. Similarly the *JPEG quality measure* (x_{23}) is an automated image quality measure, considering artifacts caused by JPEG compression, such as blockiness and blurriness, evaluating again a no-reference score per image [64].

2.3 Facial characteristics

Literature related to facial beauty (cf. [65]) identifies pertinent traits including the size of the face, the location and size of facial features like eyes, nose, and mouth, brows, lashes and lids, facial proportions, as well as the condition of the skin. Such literature confirms the role of these facial features in affecting human perception of beauty (cf. [4, 65]). Drawing from this previous work, we deduce the characteristics $x_1, x_2, x_7, x_{12}, x_{13}, x_{15}, x_{16}, x_{19}, x_{22}, x_{25}, x_{29}, x_{30}, x_{36}, x_{37}$ and also consider ratios of facial features and / or their locations by relating a multitude of measures, specifically including known facial beauty ratios adhering to the golden ratio, e.g. x_{16} (see Table 2 for notations). In addition, we include one symmetry characteristic (x_{24}), as introduced in [57], which defines the ratio between left eye- and right eye-width to be between 0.93 and 1.06 in an attractive face. We also consider the skin-goodness / quality (x_{10}) and lip fullness (x_6), as suggested and validated in [5]. Moreover we proceed a step further and consider non-permanent characteristics that carry low discriminative biometric information. Such characteristics include eye-color (x_{32}), hair-color

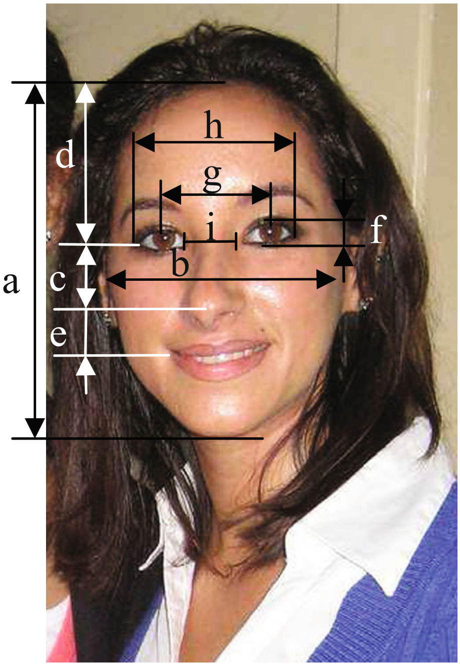

Fig. 6 Example image of the web site HOTorNOT, $MOS = 9.2$ with employed facial measures

(x_{33}) and skin-color (28), face-shape (x_4) and eye brow-shape (x_5), as well as (with a slight abuse of notation) presence of glasses (x_8), makeup style (x_3, x_9) and hair style (x_{11}). For more information on use and versatility of facial non-permanent traits see for example [15–17, 19]. The impact of makeup on face recognition has been established in [9, 11, 18].

The full set of facial features is listed in Table 2 and can be categorized in the following four groups:

- ratios of facial features and their locations,
- facial color traits,
- shapes of face and facial features,
- non-permanent traits.

Features related to the mouth and nose width were not exploited, due to the variety of expressions within the dataset. This expression variety causes significant diversity in the measurements of both, mouth and nose. All selected traits are listed in Table 2. The table exhibits the traits, trait instances and furthermore the range of magnitude for all photograph aesthetics and facial aesthetics, respectively.

3 Results

3.1 Effect of traits on the MOS rating

We here aim to find correlation measures for each of the 37 extracted traits and the MOS in order to observe the importance of each characteristic for human perception. The

Table 2 The 37 characteristics used in the proposed metric listed in decreasing order with respect to the absolute Pearson's correlation coefficient, computed between each characteristic and the averaged beauty rating (MOS)

Trait x_i	Trait instance	Pearson's correlation coefficient $r_{x_i,MOS}$	$\widehat{MOS}$-Model weight γ_i
x_1. Ratio (eye height/ head length) f/a	Continuous	0.511	18.351
x_2. Ratio (head width/ head length) b/a	Continuous	0.449	4.578
x_3. Eye make up	0: No makeup, 0.5: light makeup, 1: strong makeup	0.379	0.305
x_4. Face shape	1: round, 2: oval, 3: edgy	0.352	0.161
x_5. Eye Brow shape	1: straight, 2: round, 3: edgy	0.252	0.334
x_6. Fullness of Lips	0: Thin lips, 0.5: medium, 1: full lips	0.224	0.202
x_7. Ratio (from top of head to nose / head length) (d+c)/a	Continuous	0.220	−17.828
x_8. Glasses	0: No glasses, 1: glasses	−0.209	−0.671
x_9. Lipstick	0: No lipstick, 1: bright lipstick, 2: flashy lipstick	0.200	0.050
x_{10}. Skin goodness	1:Clear skin, 2: bad skin (e.g. pimples)	−0.186	−0.393
x_{11}. Hair Length / Style	1: Short, 2: shoulder, 3: long, 4: half tied back, 5: tied back	−0.185	−0.066
x_{12}. Ratio (from top of head to mouth / head length) (d+c+e)/a	Continuous	0.182	−4.192
x_{13}. Ratio (from top of head to eye / head length) d/a	Continuous	0.177	49.394
x_{14}. Image format	1: Portrait, 2: Landscape	0.168	0.169
x_{15}. Ratio (eye width / distance between eyes) (h-i)/(2.i)	Continuous	0.134	0.898
x_{16}. Ratio (from nose to chin / eye to nose) (a-d-c)/c	Continuous	−0.120	0.097
x_{17}. Left eye distance to middle of image or to mass point	1: shorter distance to middle of image, 2: shorter distance to mass point	0.118	0.420
x_{18}. Right eye distance to middle of image or to mass point	1: shorter distance to of image, 2: shorter distance to mass point	0.115	0.204
x_{19}. Ratio (from top of head eye / eye to nose) d/c	Continuous	−0.101	−1.009
x_{20}. Image Resolution	Normalized from 0 to 1; Discrete	0.101	−0.349
x_{21}. Expression	1: Smile + teeth, 2: smile, 3: neutral, 4: corners of the mouth facing down, 5: non of all	−0.091	−0.318

Table 2 (continued)

Trait x_i	Trait instance	Pearson's correlation coefficient $r_{x_i,MOS}$	$\widehat{MOS}$-Model weight γ_i
x_{22}. Ratio (outside distance between eyes / top of the head to eye) h/d	Continuous	−0.083	−1.726
x_{23}. JPEG quality measure	Continuous	0.080	0.901
x_{24}. Eyes symmetry	0.93 <(left eye width)/ (right eye width) < 1.06	−0.065	−0.055
x_{25}. Ratio (from eye to nose / nose to mouth) c/e	Continuous	0.064	0.046
x_{26}. Nose distance to middle of image of mass point	1: shorter distance to middle of image, 2: shorter distance to mass point	0.054	0.017
x_{27}. Illumination	0: poor; 0.5: medium; 1: excellent	0.037	0.013
x_{28}. Skin Color	1, 2, 3 (from light to dark)	−0.037	−0.055
x_{29}. Ratio (from top of head to eye / eye to lip) d/(c+e)	Continuous	0.033	−6.247
x_{30}. Ratio (eye-nose/head width) c/b	Continuous	0.025	−0.632
x_{31}. Zoomfactor a / Image resolution	Continuous	−0.020	−148.74
x_{32}. Eye Color	1: blue, 2: green, 3: brown, 4: black, 5: mix	−0.018	−0.016
x_{33}. Hair Color	1: blond, 2: brown, 3: black, 4: red, 5: dark blond	−0.017	0.031
x_{34}. Angle of face α	Continuous	−0.014	−0.269
x_{35}. BIQI	Continuous	0.012	−0.005
x_{36}. Ratio (from nose to chin / lips to chin) (a-d-c)/(a-d-c-e)	Continuous	−0.006	−1.691
x_{37}. Ratio (Distance eyes/ head length) g/a	Continuous	−0.003	13.959

Related Pearson's correlation coefficients and related MOS-model weights are reported. See Figs. 3 and 4 for the annotation of the facial characteristics

preprocessing step for the MOS related study includes the removal of about 5 % of the images, due to their outlier character (i.e. $> 2\sigma_{x_i}$, given that x_i is each function of the described traits and σ_{x_i} is the associated standard deviation).

A direct way to find a relationship between the MOS and each of the 37 traits is using Pearson's correlation coefficient. We remind the reader that for two vectors, $X = x_1, x_2, , x_n$ and $Y = y_1, y_2, , y_n$, the Pearson's correlation coefficient is given by

$$r_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}, \tag{1}$$

where σ_X and σ_Y are being the standard deviations for X and Y, respectively. The coefficient ranges between -1 and 1, with the two extreme points being obtained when the variables are maximally linearly related.

Pearson's correlation coefficients are calculated for all 37 vectors and the MOS respectively, each vector corresponding to a feature. Per feature, a 260-values X vector describes each feature for each one of the 260 training images.[2] The 260–values vector Y describes the related MOS rating per image. Table 2 itemizes these coefficients in decreasing order of importance with respect to the absolute Pearson's correlation coefficient.

3.2 Insight provided from empirical data

The first notable result reveals the strong correlation between the best ranked traits and the MOS, which even exceeds a Pearson's correlation coefficient of 0.5 for the trait 'ratio eye-height/face-height'. Particularly in regard to an automatic MOS prediction image processing tool, these results are very encouraging. Further we observe that photo-quality features play a less significant role than facial aesthetics, as expected, but the related pertinence is not to be neglected, since they achieve an $r_{14,MOS} = 0.168$. Moreover we note that the high ranked traits x_1, x_2 and x_4 (the ratios (eye-height/face-height) and (head-width/head-height), and face shape) in Table 2 are features corresponding strongly to person's weight. This outcome brings to the fore the strong importance of low human weight for aesthetics [52], that has been reported in other studies [26], [6]. Furthermore it is worth noting that Table 2 reveals the interesting fact among others, that non-permanent traits place a pivotal role in raising the MOS rating. Eye makeup, lipstick, glasses and hair-style are all among the top 11 of the obtained ranking. These results hint the high modifiability of facial aesthetics perception by simple means including makeup and hair styling. The relevance of eye makeup had been previously observed in [30]. Together with the different conclusions that one may draw from Table 2, it also becomes apparent that different questions are raised, on the interconnectedness of the different traits. This is addressed in the following Section 3.3. Finally we note that traits, such as x_1, x_7, x_{12} and x_{13} directly comply with the well known babyfaceness hypothesis (cf. [65]), which describes that childlike facial features in females increase attractiveness, such features include big eyes, cf. x_1 and a relative low location of facial elements, cf. x_7, x_{12} and x_{13}. One measure known for increasing attractiveness, if equal to the golden ratio $\phi = 1.618$, is x_{16}.

3.3 Interconnectedness of different traits

To obtain a better understanding of the role of the different traits in raising the MOS, it is helpful to understand the inter-relationship between these traits. This is addressed in Table 3, which describes the correlation between selected traits. Due to lack of space we limit the correlation matrix to just a group of the first six traits. Table 3 can answer different questions such as for example the validity of the conclusion in Table 2 on the importance of the makeup feature. In this case, the question arises whether it is truly the makeup that affects the MOS or whether already attractive subjects use makeup more heavily. Table 3 suggests a low correlation between the facial proportions (representing beauty) and eye makeup, which validates the strong role of makeup in raising the MOS.

[2] For information on denotation of features and according X–values, please refer to Table 2

Table 3 Correlation matrix of selected non-permanent and permanent traits, see Table 2

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1	0.317	0.308	0.153	0.151	0.161
x_2	0.317	1	0.132	0.268	0.034	0.092
x_3	0.308	0.132	1	0.140	0.158	0.108
x_4	0.153	0.268	0.140	1	−0.0036	0.122
x_5	0.151	0.034	0.158	−0.0036	1	0.155
x_6	0.161	0.092	0.108	0.122	0.155	1

4 Model for facial aesthetics

We here examine two different regression techniques, namely linear regression with multivariate data and support vector regression SVR.[3]

4.1 Linear regression

Linear regression benefits from its simplicity and the linear character of the traits with increasing MOS. We have the following form:

$$\widehat{MOS} = \sum_{i=1}^{37} \gamma_i x_i . \tag{2}$$

The resulting weights γ_i corresponding to each trait are denoted in Table 2.

We here note that the weights of the model are not normalized and do not give information about the importance of each characteristic. In other words, we did not normalize for the sake of reproducibility - $\widehat{MOS}$ can be computed with features labeled as in Table 2 (and related weights from the same table). The importance of the characteristics can be gleaned from the Pearson's correlation coefficients $r_{x_i,MOS}$ instead.

To test the efficacy of the scheme, first we randomly partition the dataset into a training (260 subjects) and a test set (65 subjects) and exclude outliers.

We briefly compute the following three parameters.

- Pearson's correlation coefficient is computed as described above, and it is computed to be $r_{\widehat{MOS},MOS} = 0.779$.
- Spearman's rank correlation coefficient, which is a measure of how well the relation between two variables can be described by a monotonic function. The coefficient ranges between -1 and 1, with the two extreme points being obtained when the variables are purely monotonic functions of each other. This coefficient takes the form $r_S = 1 - \frac{6\sum_i d_i}{n(n^2-1)}$, where $d_i = rank(x_i) - rank(y_i)$ is the difference between the ranks of the i^{th} observation of the two variables. The variable n denotes the number of observations. The coefficient, which is often used due to its robustness to outliers, was calculated here to be $r_{S\widehat{MOS},MOS} = 0.786$.

[3]SVR has been used in similar problems, such as automated age estimation with a reportedly high performance in mapping of actual age to automatically extracted feature vectors [28].

– Mean standard error of the difference between the estimated objective $\widehat{MOS}$ and the actual subjective MOS: $MSE = 1.158$.

The high Pearson's coefficient implies a robust prediction accuracy of the facial aesthetics metric. The Spearman's coefficient gives an indication about the correlation between estimated and real MOS, but without the restriction of linear dependence. It considers each monotonic function connecting the two vectors. In our case this coefficient is relatively high as well. The MSE on the other hand gives an idea about the absolute error between the predicted and actual values.

4.2 Support vector regression

In support vector regression the training samples $(x_1, y_1), \ldots, (x_m, y_m)$ include the inputs $x \in \mathbb{R}^m$, that are m-dimensional vectors (feature vectors) and the outcomes $y \in \mathbb{R}$ are continuous values (beauty indices). Firstly, x is mapped onto an n-dimensional feature space using a fixed nonlinear mapping, and then a linear model is constructed in this feature space. Hereby, SV - regression uses an ε - insensitive loss function, as proposed by Vapnik [59], where ε is a radius of a tube within which the regression function must lie, after the successful learning. By aiming to reduce model's complexity, SV-regression is formulated as the minimization of

$$min \frac{1}{2}\|\omega\|^2 + C \sum_{i=1}^{n} \left(\xi_i + \xi_i^T\right), \tag{3}$$

subject to the constraint:

$$\begin{cases} y_i - f(x_i, \omega) \leq \varepsilon + \xi_i^T \\ f(x_i, \omega) - y_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^T \geq 0, i = 1, ..., n. \end{cases} \tag{4}$$

SV-regression performs in the high-dimension feature space, reducing the model complexity by minimizing $\|\omega\|^2$ (capacity of machine learning). We have that ξ_i, ξ_i^T are slack variables, which measure the deviation of training samples outside ε - insensitive zone. Parameter C determines the trade off between the model complexity and the degree, to which deviations larger than ε, are tolerated in optimization formulation. Parameter ε controls the width of the ε-insensitive zone, used to fit the training data.

4.3 Validation of the obtained metric

We use the presented database, exclude outliers and adopt a 6-fold cross-validation scheme for performance evaluation of both regression techniques. To validate the proposed model we compute the Pearson's correlation coefficient, as described above. We present the results in Table 4.

It is interesting to observe that linear regression outperforms the SV-regression. We here note that for SV-regression we utilized the LIBSVM [7]-library and specifically SV-regression with linear kernel (which outperformed the here not presented types of SV-regression, namely polynomial kernel, radial basis function and sigmoid), while for linear regression we used the MATLAB functions.

In the following, we proceed to compare the presented metric (utilizing linear regression) with two reference state of the art algorithms.

Table 4 Comparison of linear and support vector regression for the proposed feature based beauty estimation metric in the presented database

Trial	Train set	Test set	lin. Regr.	SV-Regr.
1	260	38	0.768	0.727
2	238	60	0.626	0.591
3	248	50	0.68	0.559
4	248	50	0.532	0.332
5	248	50	0.671	0.602
6	248	50	0.621	0.532
Average			0.65	0.557

The numbers in the "lin. Regr." and "SV-Regr." columns indicate the Pearson's correlation coefficient (r) between estimated $\widehat{MOS}$ and actual MOS in each trial

4.4 Reference beauty prediction algorithms

In order to directly compare the proposed algorithm with already existing algorithms, we proceed to re-implement one geometric-based (=feature-based) and one holistic algorithm perform beauty estimation with these two algorithms on the presented dataset.

4.4.1 Reference feature based algorithm

We re-implement the algorithm by Mao et al. [43], where 17 geometric features are manually extracted in a human face. The features include: a) horizontal face length at temple level, b) horizontal face length at cheekbone level (ears are excluded) c) horizontal face length at cheekbone level, d) horizontal face length at mouth level, e) horizontal length of chin, f) horizontal length of nose, g) horizontal length of mouth, h) horizontal distance between pupils, i) horizontal length of left eye, j) horizontal distance between inner edges of eyes, k) horizontal length of right eye, l) vertical distance from middle point between eyebrows to the bottom of nose, m) vertical distance from nose bottom to face bottom, n) vertical distance between eyebrows and eyes, o) vertical height of eyes, p) vertical distance between nose and mouth, and q) vertical distance between mouth and chin.

Feature j is used to normalize the remaining 16 features. In the initial experiment the algorithm was evaluated on a 510-Chinese female dataset, that is not publicly available. Beauty estimation in this original paper is posed as a beauty *classification* problem, where human raters had rated the dataset into 4 categories (very attractive, attractive, low attractive, not attractive). The *reported results* [43] encompassed a classification rate of 95.3 % for 2-level classification (when classifying 2 out of 4 classes, namely the highest and the lowest) and 77.9 % for 4-level classification.

Experiments of the reference feature based algorithm on presented database: Similarly, we utilize the suggested 17 features, and subsequently use Support Vector Regression and multi-variate linear regression to estimate a continuous beauty index $\widehat{MOS}$, in order to compare the obtained correlation performance with the one achieved by the proposed algorithm. Hence, we annotate the 17 features manually in the proposed database, train then the Support Vector Regression and linear regression, respectively with the same 6-fold crossvalidation scheme as Table 4 and obtain results, that we present in Table 5.

Table 5 Comparison of three beauty estimation algorithms: (a) reference feature based [43], reference holistic [25] and the proposed feature-based algorithm

Trial #	Train set	Test set	Feature based Alg. [43] SV-regr.	Feature based Alg. [43] Lin. regr.	Holistic Alg. [25]	Proposed Alg.
1	260	38	0.560	0.601	0.500	0.768
2	238	60	0.452	0.456	0.416	0.626
3	248	50	0.597	0.632	0.41	0.68
4	248	50	0.593	0.574	0.497	0.532
5	248	50	0.520	0.485	0.243	0.671
6	248	50	0.515	0.622	0.362	0.621
Average			0.562	0.561	0.405	0.65

The numbers in the four right columns present the Pearson's correlation coefficient in each trial of the 6-fold crossvalidation experiment. Abbreviations: Alg. ... algorithm, SV-regr. ... Support Vector regression, Lin. regr. ... linear regression

4.4.2 Reference holistic algorithm

Additionally, we re-implement a fully automated algorithm based on Eigenfaces [25]. The original experiment was evaluated on a dataset containing 184 high quality images of Caucasian females in frontal view and with neutral expression, the images were captured in identical acquisition conditions (sensor, lighting, background). The beauty indices of 92 images were rated by 28 human raters, the remaining 92 were rated by 18 additional human raters on a scale from 1 (very unattractive) to 7 (very attractive). The algorithm projects the images onto eigenvectors to obtain a low-dimensional representation and trains classification- or regression-algorithms based on this representation. The *reported results* in [25] for classification, where the highest 25 % rated images and the lowest 25 % were classified by k-nearest neighbor classification and support vector machines (SVM) were 75 % − 85 % true classification rates. The same approach with SV-regression *reportedly obtained* in their experiment a correlation rate of $\rho = 0.45$.

Experiments of holistic reference algorithm on presented database: We conduct a similar experiment with Eigenfaces and SV-regression, by spanning the PCA-identity space with 300 random images of the HOTorNOT-based dataset assembled in [67]. We rotate and normalize the images of the here presented dataset based on the eye coordinates and obtain results shown in Table 5.

4.5 Discussion

We observe that the proposed algorithm significantly outperforms the other two algorithms. As expected, both reference algorithms are challenged with the presented database, due to variations in expression and illumination. The holistic algorithm is additionally challenged by different image quality and acquisition conditions. The reference feature based algorithm has a better performance than the holistic one, but is challenged mainly by different expressions, that are unwillingly considered in several features related to the nose and mouth. Both, feature-based vs. holistic algorithms have benefits and limitations, that can be outweighed based on the related application / database. In particular the nature of the utilized database is one significant factor for the performance of algorithms. In constrained datasets (limited variations in pose and expression of the subjects), geometric-based algorithms provide better results than holistic algorithms, see Table 1. The limitation of such though clearly is the time-consuming manual annotation of facial landmarks, that cannot be automated easily.

Absolute comparisons to other beauty-estimation works (e.g. from Table 1) are challenging, due to the heterogeneous nature of aspects such as datasets (constrained / unconstrained, dataset-size, image-size and quality), as well as methods of evaluation. Additionally, while some approaches classify faces into discrete beauty categories (e.g. 2 categories: attractive / non-attractive, or more refined 4 or 7 categories) and use the true classification rate to assess performance, other methods compute a continuous real-valued beauty index and utilize the correlation between the values assigned by human raters and estimated values to assess performance. Also, the rating of attractiveness differs in the number of human raters and the used beauty-scale.

We proceed with three experiments using the validated MOS–prediction metric.

5 Experiments with $\widehat{MOS}$

The proposed MOS prediction metric is here employed to quantify beauty. While different insights can be drawn, one can consider specific questions such as:

- Are famous females known for their beauty more beautiful than average females?
- What is the impact of age on beauty?
- What is the impact of facial surgery on beauty?

With the above in mind, we proceed to apply the presented metric on images drawn from the internet and from official datasets such as the FG-NET [4] and the plastic surgery dataset [56].

5.1 Metric verification on highly ranked females

Towards verification of its usefulness, we applied the above designed MOS–prediction metric on images of females who have been highly ranked for their beauty by the popular media. Specifically we considered images of females leading the lists of *People's magazine* as the 'most beautiful people' from 1991 to 2013, as well as the top 10 entries from the same list for the year 2010. The considered subjects included, among others, Jennifer Lopez, Julia Roberts and Angelina Jolie. We have assembled facial images of named subjects, annotated them and computed the related beauty indices. We opposed the obtained results with results obtained from the images of the HOTorNOT dataset. In Fig. 7 we observe the average estimated $\widehat{MOS}$ for the first dataset and average actual MOS of the HOTorNOT dataset, displaying MOS and $\widehat{MOS}$ values, as well as the associated confidence interval of 95 % confidence level. The confidence interval indicates the reliability of the computed average-MOS values. Thus there is a 95 % likelihood that the true average-MOS value for both populations (highly attractive females and HOTorNOT-females) lies in the respective confidence interval. We note, that the confidence interval depends on three factors: a) sample size, b) percentage (of a sample for a particular answer) and c) population size. A large sample size provides with more certainty the true reflect of the population. For a given confidence level (e.g. 95 %), the larger the sample size is, the smaller the confidence interval will be. In our case, the sample size of the "People's magazine most beautiful females" is

[4]FG–NET dataset http://www.fgnet.rsunit.com/

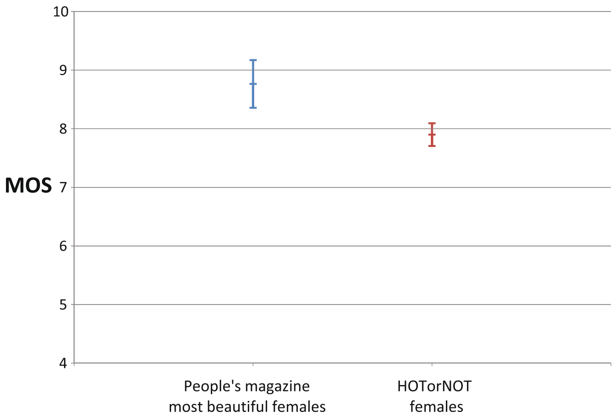

Fig. 7 Comparison of average estimated MOS for the assembled HOTorNOT dataset and the average $\widehat{MOS}$ for the People's magazine most beautiful famous females dataset

22, the sample of the "HOTorNOT-females" is 260 and thus the second confidence interval is much smaller. The test validates the presented beauty estimation metric, with the entries from the above 'beautiful people' lists consistently scoring significantly higher scores.

5.2 Intercorrelation between beauty and age: FG–NET aging dataset

Towards investigating the dependence between beauty and age, we considered images from the FG–NET dataset, as this dataset provides us with multiple images of subjects as they age. We specifically selected females, with a broad time spectrum images, in other words females with images available from an age about 18 years old to 60 years old. We annotated these images in terms of employed facial features and computed the corresponding $\widehat{MOS}$ values. We obtained per subject several beauty scores spanned over time. Since the range of these beauty functions differed on the MOS scale between females, we normalized the computed $\widehat{MOS}$ value per subject. We then averaged the normalized beauty over time functions and estimated based on the result a polynomial function of the 5th degree. Figure 8 displays the merged functions and the related estimation function. The resulting beauty function over time bares a maximum between the ages 23 to 33. The outcome can be explained by the facts, that with advanced age generally (a) the skin quality decreases (wrinkles, aging spots), (b) glasses are worn, as well as (c) the interest in regards to makeup or hair style decreases.

5.3 Facial surgery

We also examined the effect of *blepharoplasty* (eyelid lifting surgery) on the estimated beauty index. Our choice of this specific parameter and surgery was motivated by the fact that eye size has been shown to have a high impact on the presented beauty metric. We randomly selected 20 image pairs (before and after the surgery,[5] see Fig. 9) from the

[5]For this experiment all values attached to non-permanent traits were artificially kept constant for "before surgery" and "after surgery" images.

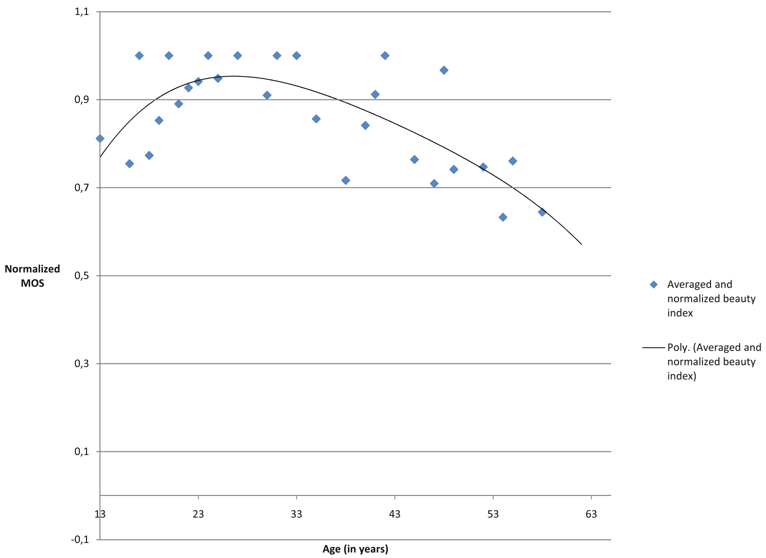

Fig. 8 Normalized $\widehat{MOS}$ values for females in the course of time

plastic surgery dataset [56], and after annotation, we computed the related beauty indices. Interestingly the presented analysis suggested a relatively small surgery gain in the $\widehat{MOS}$ increase. Specifically the increase revealed a modest surgery impact on the beauty index, with variations ranging in average between 1 % and 4 %.

6 Future work

Future work will involve the development of an automated beauty assessment tool. It is noted that such a tool, that would be based on all 37 of the aforementioned characteristics, would - in addition to a maximal prediction score - also introduce additional estimation errors into the prediction performance. For this reason, future analysis and design of such automated tools, must carefully consider the tradeoff between possible prediction score

Fig. 9 Example images of two subjects of the plastic surgery dataset [56]. For each subject the left image depicts the face *before surgery*, the right image *after surgery*. From left to right we have computed following $\widehat{MOS}$: 5.17, 5.47, 7.96, 8.8

Table 6 Limited sets of traits and associated Pearson's correlation coefficients with MOS

Trait x_i	Pearson's correlation coefficient $r_{i,MOS}$
x_1	0.5112
x_1, x_2	0.5921
x_1, x_2, x_{12}	0.5923
x_1, x_2, x_3	0.6319
x_1, x_2, x_8	0.6165
x_1, x_2, x_{12}, x_{15}	0.5930
x_1, x_2, x_3, x_8	0.6502
$x_1, x_2, x_{12}, x_{15}, x_{14}$	0.6070
$x_1, x_2, x_4, x_{12}, x_{14}, x_{15}$	0.6392
$x_1, x_2, x_4, x_5, x_{12}, x_{14}, x_{15}$	0.6662
$x_1, x_2, x_4, x_5, x_{12}, x_{14}, x_{15}, x_{20}$	0.6711
$x_1, x_2, x_8, x_{14}, x_{20}, x_{23}$	0.6357

We have selected these sets by both, their pertinence, as well as easiness for automated extraction: e.g. limited characteristic facial landmarks and glasses detection. We note that the features x_{14}, x_{20}, x_{23}, namely image format, JPEG quality measure and image resolution are already automatically obtained

and classification error. We illustrate a preliminary analysis of prediction scores evoked by different combinations of small sets of traits in Table 6.

We observe that a limited set of 8 traits can achieve a relatively high Pearson's correlation coefficient of 0.67. An optimization towards more photo-quality related traits obtains 0.6358. This result motivates future work based on few facial landmarks (as shown in [22]), face localization (see [33]) and glasses detection (see [68]). The features x_{14}, x_{20}, x_{23}, namely image format, JPEG quality measure and image resolution are already automatically obtained.

7 Conclusions

In this work, we presented a study on facial aesthetics in photographs, where we explored the matching of objective measures (namely photograph quality measures, facial beauty characteristics and non-permanent facial features), with human subjective perception. The presented analysis revealed a substantial correlation between different selected traits, and the corresponding MOS-related beauty indices. Specifically we provided evidence that non-permanent features can significantly influence the MOS, and based on our analysis we conclude that facial aesthetics in images can indeed be substantially modifiable. In other words, parameters such as the presence of makeup and glasses, image quality, as well as different image post–processing methods can significantly affect the resulting MOS.

Furthermore we constructed a MOS–based metric which outperformed two reference state of the art algorithms (one feature-based and one holistic), that we compared it with. Additionally, we provided an analysis that quantifies beauty-index variations due to aging and surgery. Our work applies towards building a basis for designing new image-processing tools that further automate prediction of aesthetics in facial images. Towards this we provided insight on facial aesthetics prediction.

References

1. Aarabi P, Hughes D, Mohajer K, Emami M (2001) The automatic measurement of facial beauty. In: Proceedings of IEEE SMC
2. Alley T, Cunningham M (1991) Averaged faces are attractive, but very attractive faces are not average. Psychol Sci 2:123–125
3. Bhattacharya S, Sukthankar R, Shah M (2010) A framework for photo-quality assessment and enhancement based on visual aesthetics. In: Procedings of ACM MM
4. Bottino A, Laurentini A (2010) The analysis of facial beauty: an ermerging area of research in pattern analysis. Lect Notes Comput Sci Image Anal Recogn:6111
5. Braun C, Gruendl M, Marberger C, Scherber C (2001) Beautycheck - ursachen und folgen von attraktivitaet
6. Buggio L, Vercellini P, Somigliana E, Vigano P, Frattaruolo MP, Fedele L (2012) "you are so beautiful"*: Behind women's attractiveness towards the biology of reproduction: a narrative review. Gynecol Endocrinol 28(10):753–757
7. Chang C-C, Lin C-J (2011) LIBSVM : a library for support vector machines. ACM Trans Int Syst Tech 2(27)
8. Chang F, Chou C-H (2009) A bi-prototype theory of facial attractiveness. Neural Comput 21(3):890–910
9. Chen C, Dantcheva A, Ross A (2013) Automatic facial makeup detection with application in face recognition. In: 2013 international conference on biometrics (ICB), pp 1–8
10. Chen F, Zhang D (2010) A benchmark for geometric facial beauty study. Medical Biometrics 6165:21–32
11. Chen C, Dantcheva A, Ross A (2014) Impact of facial cosmetics on automatic gender and age estimation algorithms. VISAPP'14
12. Cheng B, Ni S, Yan B, Tian Q (2010) Learning to photograph. In: Proceedings of ACM MM
13. Chiang W-C, Lin H-H, Huang C-S, Lo L-J, Wan S-Y (2014) The cluster assessment of facial attractiveness using fuzzy neural network classifier based on 3d moire features. Pattern Recogn 47(3):1249–1260
14. Dantcheva A, Dugelay J-L (2011) Female facial aesthetics based on soft biometrics and photo-quality. In: Proceedings of ICME
15. Dantcheva A, Dugelay J-L, Elia P (2010) Person recognition using a bag of facial soft biometrics (bofsb). In: 2010 IEEE international workshop on multimedia signal processing (MMSP), pp 511–516
16. Dantcheva A, Dugelay J, Elia P (2010) Soft biometrics systems: Reliability and asymptotic bounds. In: 2010 4th IEEE international conference on biometrics: theory applications and systems (BTAS), pp 1–6
17. Dantcheva A, Dugelay J-L (2011) Frontal-to-side face re-identification based on hair, skin and clothes patches. In: 2011 8th IEEE international conference on advanced video and signal-based surveillance (AVSS), pp 309–313
18. Dantcheva A, Chen C, Ross A (2012) Can facial cosmetics affect the matching accuracy of face recognition systems? In: IEEE international conference on biometrics: theory, applications and systems
19. Dantcheva A, Singh A, Elia P, Dugelay J-L (2011) Search pruning in video surveillance systems: efficiency-reliability tradeoff. In: Proceedings of international conference on computer vision workshops
20. (2013). Dating and rating webpage HOTorNOT webpage
21. Davis BC, Lazebnik S (2008) Analysis of human attractiveness using manifold kernel regression. In: Proceedings of IEEE ICIP
22. Ding L, Martinez AM (2008) Precise detailed detection of faces and facial features. In: Proceedings of CVPR
23. Doczi G (2005) The power of limits: proportional harmonies in nature, art, and architecture. Shambhala Publications, Boston
24. Eisenthal Y, Dror G, Ruppin E (2006) Facial attractiveness: beauty and the machine. Neural Comput 18(1):119–142
25. Eisental Y, Dror G, Leyvand T, Cohen-Or D, Ruppin E (2006) Facial attractiveness: beauty and the machine. Neural Comput 18(1):119–142
26. Fisher ML, Voracek M (2006) The shape of beauty: determinants of female physical attractiveness. J Cosmet Dermatol 5:190–194
27. Fink B, Grammer K, Matts PJ (2006) Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. Evol Hum Behav 27:433–442
28. Fu Y, Guo G-D, Huang T (2010) Age synthesis and estimation via faces: a survey. IEEE Trans Pattern Anal Mach Intell 32(11):1955–1976

29. Gao W, Cao B, Shan S, Chen X, Zhou D, Zhang X, Zhao D (2008) The cas-peal large-scale chinese face database and baseline evaluations. IEEE Trans Syst Man Cybernet (Part A) 38(1):149–161
30. Gray D, Yu W. XuK, Gong Y (2010) Predicting facial beauty without landmarks. In: Proceedings of ECCV
31. Gunes H (2011) A survey of perception and computation of human beauty. In: Proceedings of J-HGBU ’11 joint ACM workshop on Human gesture and behavior understanding, pp 19–24
32. Gunes H, Piccardi M, Jan T (2004) Comparative beauty classification for pre-surgery planning. In: Proceedings of SMC
33. Guo J-M, Lin C-C, Wu M-F, Chang C-H, Lee H (2011) Complexity reduced face detection using probability-based face mask prefiltering and pixel-based hierarchical-feature adaboosting. IEEE Signal Processing Letters
34. Irem Türkmen H, Kurt Z, Karsligil ME (2007) Global feature based female facial beauty decision system. In: Proceedings of EUSIPCO
35. Jonas P (1976) Photographic composition simplified. Amphoto
36. Ji HI, Kamachi M, Akamatsu S (2004) Analyses of facial attractiveness on feminised and juvenilised faces. Perception 33:135145
37. Kagian A, Dror G, Leyvand T, Cohen-Or D, Ruppin E (2006) A humanlike predictor of facial attractiveness. In: Proceedings of NIPS, no 19, pp 674–683
38. Kagian A, Dror G, Leyvand T, Meilijson I, Cohen-Or D, Ruppin E (2008) A machine learning predictor of facial attracitveness revealing human-like psychophysical biases. Vis Res 48(2):235–243
39. Kendall MG, Stuart A (1973) The advanced theory of statistics: inference and relationship, vol 2. Griffin
40. Langlois JH, Roggman LA (1990) Attractive faces are only average. Psychol Sci 1:115–121
41. Laurentini A, Bottino A (2014) Computer analysis of face beauty: a survey. Computer Vision and Image Understanding
42. Lee L, Loewenstein G, Ariely D, Hong J, Young J (2008) If I’m not hot, are you hot or not? Physical attractiveness evaluations and dating preferences as a function of one’s own attractiveness. Psychol Sci 19(7):669–677
43. Mao H, Jin L, Du M (2009) Automatic classification of chinese female facial beauty using support vector machine. In: Proceedings of IEEE SMC
44. McKeen S (2006) A beauty fix plumps up psyche and overall health. The Edmonton Journal
45. Mizumoto Y, Deguchi T, Fong KW (2009) Assessment of facial golden proportions among young japanese women. American journal of orthodontics and dentofacial orthopedics. Image Vis Comput 136:168–174
46. Moorthy A, Obrador P, Oliver N (2010) Towards computational models of the visual aesthetic appeal of consumer videos. In: Proceedings of ECCV
47. Moorthy AK, Bovik AC (2009) A modular framework for constructing blind universal quality indices. IEEE Signal Processing Letters
48. Moorthy AK, Bovik AC (2009) BIQI software release
49. Obrador P, Schmidt-Hackenberg L, Oliver N (2010) The role of image composition in image aesthetics. In: Proceedings of IEEE ICIP, pp 3185–3188
50. O’Toole AJ, Price T, Vetter T, Bartlett JC, Blanz V (1999) 3D shape an d2D surface textures of human faces: The role of “averages” in attractiveness and age. Image Vision Comput 18:9–19
51. Parris C, Robinson J (1999) The bold and the beautiful according to plastic surgeons. Technical report, Dallas
52. Perrett DI, May KA, Yoshikawa S (1994) Facial shape and judgements of female attractiveness. Nature 368:239–242
53. Said CP, Todorov A (2011) A statistical model of facial attractiveness. Psychol Sci 22 (9):1183–1190
54. Savakis A, Etz S, Loui A (2000) Evaluation of image appeal in consumer photography. In: Proceedings of SPIE, pp 111–121
55. Sheikh HR, Sabir MF, Bovik AC (2006) A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans Image Process 15(11):3440–3451
56. Singh R, Vatsa M, Bhatt HS, Bharadwaj S, Noore A, Nooreyezdan SS (2010) Plastic surgery: a new dimension to face recognition. IEEE Trans Inf Foren Sec 5(3):441–448
57. Sutic D, Breskovic I, Huic R, Jukic I (2010) Automatic evaluation of facial attractiveness. In: Proceedings of MIPRO
58. (2014) Universal concepts of beauty. Prime, vol 4
59. Vapnik V (2006) Estimation of Dependences Based on Empirical Data: Empirical Inference Science. Information Science and Statistics. Springer, Dordrecht

60. Voak I, Perrett D, Peirce J (1999) Computer graphic studies of the role of facial similarity in judgements of attractiveness. Curr Psychol 18(1):104-117
61. Wang J, Gong Y, Gray D (2014) Female facial beauty attribute recognition and editing. In: Fu Y (ed) Human-Centered Social Media Analytics, pp 133–148. Springer International Publishing
62. Wang Y, Dai Q, Feng R, Jiang Y-G (2013) Beauty is here: Evaluating aesthetics in videos using multimodal features and free training data. In: Proceedings of the 21st ACM International Conference on Multimedia, MM '13, pages 369–372. ACM
63. Wang Z, Bovik AC, Sheih HR, Simoncelli EP (2004) Image qulaity assessment: From error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
64. Wang Z, Sheikh HR, Bovik AC (2002) No-reference perceptual quality assessment of jpeg compressed images. In: Proceedings of IEEE ICIP
65. (2012). Webpage on beautycheck. www.beautycheck.de
66. Whitehill J, Movellan JR (2008) Personalized facial attractiveness prediction. In: Proceedings of IEEE FRG
67. White R, Eden A, Maire M (2004) Automatic prediction of human attractiveness. uc berkely technical report
68. Wu B, Ai H, Liu R (2004) Glasses detection by boosting simple wavelet features. In: Proceedings of ICPR
69. Zhang D, Zhao Q, Chen F (2011) Quantitative analysis of human facial beauty using geometric features. Pattern Recog 44(4):940–950

Antitza Dantcheva is currently a Marie Curie fellow post-doctoral researcher at the STARS team at INRIA Méditerranée in France. Previously she has been a Post-doctoral fellow at the Michigan State University and the West Virginia University in the USA. She received her PhD in Signal and Image Processing in 2011 from Telecom ParisTech, France. Antitza Dantcheva was the recipient of Best Tabula Rasa Spoofing Attack Award 2013, Best Poster Award in IAPR ICB 2013 and the Best Presentation Award in ICME 2011.

Her research interests are in computer vision, image processing and specifically facial analysis for security, health and aesthetics. She has been PI and co-PI of two biometrics-related projects, is in the Education Committee of the IEEE Biometrics Council, was coorganizer of the ECCV12 workshop Whats in a Face?, was Media co-chair at BTAS 2013, and has served as reviewer for IEEE TPAMI, IEEE TIFS, IEEE TCSVT, CVIU, PRL and many technical program committees.

Jean-Luc Dugelay received his PhD in Information Technology from the University of Rennes in 1992. His thesis work was undertaken at CCETT (France Télécom Research) at Rennes between 1989 and 1992. He then joined EURECOM in Sophia Antipolis, where he is now an Associate Professor in the Department of Multimedia Communications. His current work focuses in the domain of multimedia image processing, in particular activities in security (image forensics, biometrics and video surveillance, mini drones), and facial image processing. He has authored or coauthored over 250 publications in journals and conference proceedings, 1 book on "3D Object Processing: Compression, Indexing and Watermarking" by Wiley, 2008, 5 book chapters and 3 international patents. His research group is involved in several national projects and European projects. Prof. Dugelay is a Fellow of the IEEE. He has delivered several tutorials on digital watermarking, biometrics and compression at major international conferences, such as ACM Multimedia and IEEE ICASSP. He participated in numerous scientific events as a member of the scientific technical committees, invited speakers or session chair. He is an elected member of the EURASIP Board of Governors. He is/was Associate Editor of several international journals, such as IEEE Transactions on Image Processing and IEEE Transactions on Multimedia; Founding Editor-in-Chief, EURASIP Journal on Image and Video Processing (SpringerOpen); Member, SPS Image, Video, and Multidimensional Signal Processing Technical Committee (2002-07), SPS Multimedia Technical Committee (1999-2003 and 2005-08), and SPS Information Forensics and Security Technical Committee (2010-12); Coauthor of several conference articles that received an IEEE award in 2011, 2012 and 2013; Coorganizer, Workshop on Multimedia Signal Processing (Cannes, 2001) and the 2003 Multimodal User Authentication (Santa Barbara, 2003). In 2015, he will serve as General Co-Chair, IEEE ICIP (Québec City) and EURASIP EUSIPCO (Nice).