# "Faved!" Biometrics: Tell Me Which Image You Like and I'll Tell You Who You Are

Pietro Lovato, Manuele Bicego, *Member, IEEE,* Cristina Segalin, Alessandro Perina, Nicu Sebe, *Senior Member, IEEE,* Marco Cristani, *Member, IEEE*

***Abstract*—This paper builds upon the belief that every human being has a built-in image aesthetic evaluation system. This sort of "personal aesthetics" mostly follows certain aesthetic rules widely studied in image aesthetics (e.g., rules of thirds, colorfulness, etc.) though it likely contains some innate, unique preferences. This paper is a proof of concept of this intuition, presenting personal aesthetics as a novel behavioral biometrical trait. In our scenario, personal aesthetics activate when an individual is presented with a set of photos he may like or dislike: the goal is to distill and encode the uniqueness of his visual preferences into a compact template. To this aim, we extract a pool of low- and high-level state-of-the-art image features from a set of Flickr images preferred by a user, feeding them successively into a LASSO regressor. LASSO highlights the most discriminant cues for the individual, allowing authentication and recognition tasks. The results are surprising: given only 1 image as test, we can match the user identity against a gallery of 200 individuals definitely much better than chance; using 20 images (all preferred by a single user) as a biometrical trait, we reach an AUC of $96\%$, considering the Cumulative Matching Characteristic curve. Extensive experiments also support the interpretability of our approach, effectively modelling what is the "what we like" which distinguishes us from the others.**

***Index Terms*—personal aesthetics, image preferences, behavioral biometrics, computational aesthetics**

## I. Introduction

In the last two decades, the study and the development of biometric systems have become of paramount importance, from both a scientific and a practical point of view [1], [2]. Several biometrical traits have been designed, each analyzed from different perspectives like accuracy, efficiency, usability, acceptability, etc. From a very general point of view, they can be divided in two main classes:

- *physical / physiological* biometrical traits, encoding a physical characteristic of a person: the face, the fingerprint, the iris [3], [4], [5] – to cite the most striking examples – or even EEG signals, footprints, ears, dental configurations, and many others [6], [7], [8].
- *behavioral* biometrical traits: more than a physical feature, such traits encode a characteristic linked to the behavior of a person [9], like the gait or the signature [10], [11], [12], [13].

P. Lovato, M. Bicego, C. Segalin and M. Cristani are with the Dipartimento di Informatica, Univ. di Verona, Strada le grazie 15, 37134 Verona (Italy). M. Cristani is also with the Istituto Italiano di Tecnologia (IIT), via Morego 30, 16163 Genova (Italy). A. Perina is with Microsoft Research, One Microsoft way, Redmond (WA). N. Sebe is with the University of Trento, via Sommarive 5, 38123 Povo - Trento (Italy).
Corresponding author: Marco Cristani – Tel: +39 0458027988 – email: marco.cristani@univr.it

Among the behavioral approaches, some – the so-called HCI-based behavioral biometrics [9] – are based on the idea that every person has a unique way to interact with a personal computer: for example some methods successfully investigated the possibility of characterizing a person on the basis of keystrokes or mouse dynamics [14], [15]. In the same context, very recently some other approaches investigated the exploitation of Internet-based biometrical traits, like browsing histories [16] or chatting [17].

This paper makes a further step along this direction, and proposes a novel biometrical trait which exploits the "personal aesthetics" traits of people, i.e. those visual preferences that distinguish people from each other. Actually, it is known that people often get enjoyment from observing images and express preferences for some pictures over others. There is no scientifically comprehensive theory that explains what psychologically defines such preferences [18], even if some guidelines have been produced which suggest principles of general gratification [19], [20], [21], [22], [23], [24] – some of them have been modeled computationally in the field of Computational Media Aesthetics (CMA) [25]. For example, considering colors, a study reported in [24] showed that human subjects prefer blue and dislike yellow, unveiling intriguing continuity between animal and human color aesthetics. Regarding shape, the most important principle discussed in the literature is that of the "Golden Ratio": the idea is that a rectangle whose ratio between height and width is the same as the ratio of their sum to their maximum is more attractive than other rectangles. Recent studies limited the strength of this belief [26].

In this context, many CMA applications have been developed: from aesthetic photo ranking [27], [28] and preference-aware view recommendation systems [29], to picture quality analysis [30], [31]. Nevertheless, these technologies seem to forget the essential role that factors internal to the observer may have on preference, summarized by the old adage "beauty is in the eye of the beholder". Recent studies have shown that preference formation is a result of the interplay between subjective novelty, e.g. how new a visual stimulus seems to an observer, and how well the observer is able to extract the sense of a stimulus and to relate it to previous knowledge, defined as interpretability [32].

This paper is aimed at investigating how *identifiable* these aesthetics traits are, namely if it is possible to model the visual preferences of an individual in a unique way. To do that, a biometric recognition/authentication system is built: in the enrollment phase, the "preference model" of a user is learnt

Fig. 1. Some samples of favorite images taken at random from a Flickr user.

from a set of *preferred images*; in the verification/recognition phase, such model is tested with an unseen set of favorites preferred by a probe subject. More in detail, we take a crowdsearch approach [33] and we focus on Flickr[1], a popular website where every user can select his preferred photos, by tagging them as "favorites". This creates, for every user, a set of favorite photos, which is often very heterogeneous and whose modeling/recognition goes beyond standard computer vision tasks such as object/scene recognition (see Fig. 1 for an example). In order to infer the *personal* aesthetics trait of a given subject, we analyze his "favorites set": we characterize each image with different features, ranging from low-level color/edge statistics up to more high-level and semantic descriptors such as object detectors and overall scene statistics. LASSO regression is then exploited to learn the most discriminative aesthetic attributes, i.e., the aspects a user likes that distinguish her/him from the rest of the community: such aspects represent the template. In the experiments, involving both verification and identification, we will show that personal tastes act like a blueprint for a user, allowing to recognize him against a set of 200 users with high accuracy; in particular, given just one image from an unknown user, his identity is recognized better than with a random classifier, and this dramatically raises when considering a higher number of images.

The rest of the paper is organized as follows: the approach is detailed in Sec. II, focusing both on the employed features and on the learning strategy. Experiments on recognition and authentication are reported in Sec. III, together with an explorative analysis on how the features build the personal aesthetics. The paper ends in Sec. IV with some remarks and future perspectives on the usage of personal aesthetics traits for biometrics. Please note that this work extends substantially [34], adopting a different pool of image features and customizing the framework for biometrics purposes.

[1]http://www.flickr.com/

## II. The proposed approach

This section describes the main ingredients of our approach. In particular we will first describe how features are extracted from the images; then, the learning of the user specific preference model is detailed. Finally, the matching score computation is determined.

### A. Feature extraction

We adopted a wide, though not exhaustive, spectrum of features, here grouped into two families (see Table I). On one side, we considered the cues that focus on aesthetic aspects [28], [35], which we will refer to in the remainder as *perceptual* features: the reason is that the Flickr corpus is composed by pictures posted as "*favorite*", i.e. likely to represent the aesthetic and visual preferences of the users under examination. On the other side, we focus on the content of the images; to this end, we employed robust probabilistic object detectors [36] (for a complete list of all detectable objects see [36]); we also retained the average object area (the algorithm gives also the bounding box of the detected objects). In addition, we focused on the faces, adopting the standard Viola-Jones face detection algorithm [37] implemented in the OpenCV library. Finally, we adopted the GIST scene descriptors [38], which amounts to applying a set of oriented band-pass filters.

In the following, a short description is given for all the employed features; it is worth noting that each feature extracted in the proposed approach indicates the level of presence of a particular cue, i.e. an intensity count. This is needed for the adopted learning and matching framework: therefore, some of the standard features have been ignored, since they model categorical data, or codify measures which depend on angular quantities.

- **Use of light** A fundamental property for image aesthetics: underexposed or overexposed pictures are usually considered bad. In the HSV color space, we measured the light as the average intensity of the V channel, as in [28].

- **HSV statistics** We collected a set of features considering some simple statistical quantities over the HSV channels, namely the mean of the S channel and the standard deviation of S and V channels [35].

- **Emotion based** Saturation and Brightness can have direct influence on pleasure, arousal, and dominance, (which are the three axes of the emotion space [28], [35]), and they have been computed according to the equations:
  Pleasure = $0.69V + 0.22S$
  Arousal = $-0.31V + 0.60S$
  Dominance = $0.76V + 0.32S$
  $V$ and $S$ represent the matrices of the respective channels, multiplied by the coefficients reported above and then mediated over the pixels to get a single value per channel per image.

| Category | Name | L | Short Description |
|---|---|---|---|
| Perceptual | Use of light | 1 | Average pixel intensity of V channel [28] |
| | HSV statistics | 3 | Mean of S channel and standard deviation of S, V channels [35] |
| | Emotion-based | 3 | Amount of *Pleasure*, *Arousal*, *Dominance* [35], [39] |
| | Circular Variance | 1 | *Circular variance* of the H channel in the IHLS color space [40] |
| | Colorfulness | 1 | Colorfulness measure based on Earth Mover's Distance (EMD) [28], [35] |
| | Color Name | 11 | Amount of *Black*, *Blue*, *Brown*, *Green*, *Gray*, *Orange*, *Pink*, *Purple*, *Red*, *White*, *Yellow* [35] |
| | Entropy | 1 | Image entropy [34] |
| | Wavelet textures | 12 | Level of spatial graininess measured with a three-level (L1,L2,L3) Daubechies wavelet transform on the HSV channels [28] |
| | Tamura | 3 | Amount of *Coarseness*, *Contrast*, *Directionality* [41] |
| | GLCM-features | 12 | Amount of *Contrast*, *Correlation*, *Energy*, *Homogeneity* for each HSV channel [35] |
| | Edges | 1 | Total number of edge points, extracted with Canny [34] |
| | Level of detail | 1 | Number of regions (after mean shift segmentation) [42], [43] |
| | Regions | 1 | Average *size* of the regions (after mean shift segmentation) [42], [43] |
| | Low depth of field (DOF) | 3 | Amount of focus sharpness in the inner part of the image w.r.t. the overall focus [28], [35] |
| | Rule of thirds | 2 | Mean of S,V channels in the inner rectangle of the image [28], [35] |
| | Image parameters | 1 | Aspect ratio of the image [28], [34] |
| Content | Objects | 28 | Objects detectors [36]: we kept the number of instances and their average bounding box *size* |
| | Faces | 2 | Number and *size* of faces after Viola-Jones face detection algorithm [37] |
| | GIST descriptors | 24 | Level of openness, ruggedness, roughness and expansion for scene recognition [38]. |

TABLE I
SUMMARY OF ALL FEATURES. THE COLUMN 'L' INDICATES THE FEATURE VECTOR LENGTH FOR EACH TYPE OF FEATURE.

- **Hue Circular Variance** From the hue channel in the IHLS color space (see [40] for a detailed explanation), we extracted the *circular variance*:

$$\begin{aligned} A &= \sum_{i=1}^{N} \cos H_i, \quad B = \sum_{i=1}^{N} \sin H_i \\ R &= 1 - \frac{1}{N}\sqrt{A^2 + B^2} \end{aligned}$$

  with $H$ denoting the matrix of the hue channel and $N$ the total number of pixels in the image.

- **Colorfulness** It allows to distinguish multi-colored images from monochromatic, sepia or simply low contrast images. It is measured using the Earth Mover's Distance (EMD) between the histogram of the image and a flat histogram representing a uniform color distribution, according to the algorithm suggested by Datta *et al.* [28].

- **Color name** Each color is used in many ways by photographers, accounting for their personal style. Following [35], we considered the following 11 color names: black, blue, brown, grey, green, orange, pink, purple, red, white and yellow. Each image has been converted to HSV color space: in addition to hue, saturation and brightness values have been considered, as proposed by [44]. In practice, we count the number of pixels falling in the ranges of HSV values corresponding to a specific color name.

- **Entropy** We calculated the entropy, a statistical measure that characterizes the homogeneousness of an image.

- **Wavelet textures** They are used to measure spatial smoothness/graininess in images using the Daubechies wavelet transform as presented in [28], [35]. In practice, we computed a three-level wavelet transform on all three HSV channels. The three levels of wavelet bands are arranged from top left to bottom right, in the transformed image, and the four coefficients per level are $LL, LH, HL$, and $HH$. Denoting the coefficients (we do not take into account the $LL$ coefficient) in level $i$ for the wavelet transform of one channel of an image as $w_i^h$, $w_i^v$ and $w_i^d$, with $i = 1, 2, 3$, and $h = HH$, $v = HL$ and $d = LH$, the wavelet features are defined as: $wf_i = \frac{\sum_{x,y} w_i^h(x,y) + \sum_{x,y} w_i^v(x,y) + \sum_{x,y} w_i^d(x,y)}{(|w_i^h| + |w_i^v| + |w_i^d|}$. This is computed for every level $i$ and every channel of the image, thus we get 9 features. The values $x, y$ span over the spatial domain of the single $w$ taken into account, and the operator $|\cdot|$ accounts for the spatial area of the single $w$. The corresponding wavelet features of saturation and intensity images are computed similarly. We extract three more features by computing the sum of the average wavelet coefficients over all three frequency level for each HSV channel.

- **Tamura** These textural features have been developed to correspond to human visual perception. As in [35], we kept three of the six available features, namely coarseness, contrast, and directionality, since they attained very successful results in mimicking human perceptive mechanisms.

- **Gray-level Co-occurrence Matrix (GCLM) features** GLCM features are an additional way to specify textural properties of an image. By means of the GCLM we computed contrast, correlation, energy and homogeneity of each channel of the image converted in HSV color space as in [45].

- **Edges** We focused on the presence or absence of edges, computed with the Canny edge detector. We considered the number of edge points; in order to avoid the dependence from the possible different sizes of images, the number of edge's pixels has been normalized by the total image area.

- **Level of detail and regions** As shown in the recent work of [46], [47], objects and scene semantics are very important to understand the subjective judgement

of a picture. Following this, we performed image segmentation collecting some low-order statistics. We employed the mean shift segmentation algorithm [42], and in particular the EDISON implementation [43]. After segmenting an image we extracted *i)* the number of segments - measuring the regions "density" which characterizes each image, we can interpret this feature as the Level of Detail (an image with much detail generally produce a different psychological effect than minimalist composition) - and *ii)* the average extension of the regions. All the values have been normalized w.r.t. the total image area.

- **Low depth of field (DOF)** It corresponds to the range of distances from a camera for which a photo is acceptably sharp [28], [35]. It is used by professional photographers to blur the background, drawing the attention of the observer to the object of interest which is sharp. To detect low DOF and macro images we computed a ratio of the wavelet coefficients in the high frequency (level 3) of the inner part of the image against the whole image. We divided the image into 16 equal rectangular blocks $M1, \dots M16$, numbered in row-major order. Let $w_3 = w_3^{LH}, w_3^{HL}, w_3^{HH}$ denote the set of wavelet coefficients in the high frequency of the hue image $I_H$. The low depth of field indicator feature $f_H$ for hue is computed as follows, $f_H = \frac{\sum_{(x,y)\in M_6\cup M_7\cup M_{10}\cup M_{11}} w_3(x,y)}{\sum_{i=1}^{16}\sum_{(x,y)\in M_i} w_3(x,y)}$, with $f_S$ and $f_V$ being computed similarly for $I_S$ and $I_V$ respectively.

- **The rule of thirds** The rule of thirds in photography refers to the locations of the picture where the most interesting visual object is expected to be. Such locations are four points, which represent the intersections of four orthogonal lines. Such lines are obtained by equally dividing each size of the image in three parts, and connecting the opposite points of subdivision. The rule of thirds in computational aesthetics is an approximation of this criterion, supposing that the object of interest is stretched from an intersection up to the center of the image, and is obtained by averaging the HSV color values following the formula $\frac{9}{XY}\sum_{x=X/3}^{2X/3}\sum_{y=Y/3}^{2Y/3} I_H(x,y)$, and similarly for the other two color channels, with $X$ and $Y$ indicating the horizontal and vertical size of the image [28], [35].

- **Object Detection** Motivated by [46], [47], we employed the Deformable Part Models [48], [49] system to detect objects. The algorithm works by detecting and localizing a specific object (for example a plane, a cat, a chair or a person), through the use of a model learned from a set of training examples. The system can detect different objects; in our approach we retained the number of times every detectable object is present in the image (for a complete list of all detectable objects see [47]); we also retained the average area (the algorithm gives also the bounding box of the detected objects), to estimate if objects are more towards the background or the foreground. We excluded boats, chairs, cows, sheeps, sofas and tables objects as in the training set they were never detected.

- **Faces** As a particular class of objects - which detection has been largely studied in the field of biometrics - we extracted the number and size of the faces present in the image. We employed the standard Viola-Jones face detection algorithm [50] implemented in the OpenCV library.

- **Scenes.** Finally, we focused on describing the semantics of the whole scene, rather than the semantics of single objects which appears in it. A very powerful scene descriptor is the GIST [38], which, roughly speaking, measures the responses of different Gabor filters. Such filters are built to describe the category of the scene in terms of openness, ruggedness, roughness and expansion[2].

The concatenation of all these descriptors, a vector $\mathbf{x}_m$ of 111 elements, represents the proposed signature for the image $m$. Since every feature has a very heterogeneous range of values, each feature/dimension is normalized across the images to have zero mean and unit standard deviation. More details are given in the experimental section.

It is worth noting that for the sake of reproducibility, every parameter of the different off-the-shelf computer vision libraries has been left as the default setting.

### *B. Learning the preference model*

The preference model for the user $i$, is built starting from a set of $N$ favorite images (that is, their $D$-dimensional feature vectors) $\mathbf{x}_1^{(i)}, \dots, \mathbf{x}_N^{(i)}$, representing the biometrical trait $X^{(i)} = \{\mathbf{x}_1^{(i)}, \dots, \mathbf{x}_N^{(i)}\}$ In our case, $D = 111$.

Given the biometrical trait $X^{(i)}$, we can build the template as follows. First, we partitioned his favorite images in two sets, one for the training, $X_{\rm tr}^{(i)}$, composed by $N_{\rm tr}$ images, and one for the testing, $X_{\rm te}^{(i)}$, composed by $N_{\rm te}$ images. We will see in the next section how crucial is the choice of the value of $N_{\rm tr}$ and $N_{\rm te}$. In the following, we will also say that the images $X_{\rm tr}^{(i)}$ will define the *gallery* biometrical trait for the user $i$, while the images $X_{\rm te}^{(i)}$ will define the *probe* biometrical trait for the user $i$.

Since we are interested in characterizing which are the particular personal tastes of the given user, we decided to train a binary classifier, using as positive examples $X_{\rm tr}^{(i)}$ and as negative the favorites of other Flickr users $\{X_{\rm tr}^{(j)}\}_{j\neq i}$: this will permit to extract what really makes the subject different from the others. In particular, we represent the discriminative aesthetical aspects of each user as a subset of all the features considered, opportunely weighted. To do that, we perform a sparse regression analysis using Lasso [51]. Lasso is a general

[2]The code is publicly available on http://people.csail.mit.edu/torralba/code/spatialenvelope/

form of regularization in a regression problem. In the simple linear regression problem, every $n$-th training image, described by the proposed feature vector $\mathbf{x}_n$, is associated with a target variable $y_n$ (a positive label is given to all training images coming from user $i$, that is, the one we want to characterize, whereas the favorites of other users $j$, $j \neq i$, have a negative label). Then, we can express the target variable as a linear combination of the image features:

$$y_n = \mathbf{w}^{(i)T}\mathbf{x}_n \tag{1}$$

The standard least square estimate calculates the $D$-dimensional weight vector $\mathbf{w}^{(i)} =$ by minimizing the error function

$$E(\mathbf{w}^{(i)}) = \sum_{n=1}^{N_{\mathrm{TR}}} \left(y_n - \mathbf{w}^{(i)T}\mathbf{x}_n\right)^2 \tag{2}$$

where in our case $N_{\mathrm{TR}}$ corresponds to the total number of images of all the users we have in the training set. The regularizer in the Lasso estimate is simply expressed as a threshold on the L1-norm of the entries $\{w_d\}_{d=1,\ldots,D}$ of the weight $\mathbf{w}$:

$$\sum_{d=1}^{D} |w_d| \leq t \tag{3}$$

This term acts as a constraint that has to be taken into account when minimizing the error function.

By doing so, it has been proved that (depending on the parameter $t$), many of the coefficients $w_d$ become exactly zero [51]. Since each component $w_d$ of the weight vector weighs a different feature, it is possible to understand which features are the most important for a given user, and which ones are neglected. By looking at the values in the "user-specific" weight vector $\mathbf{w}^{(i)}$ for user $i$, we have that only the most important image features that characterize the preferences of that user are retained. Therefore, we can call $\mathbf{w}^{(i)}$ the *template* for user $i$.

More in detail, a positive weight for a feature indicates that in the pool of preferred images of a user that feature is present, and is discriminative for the user. Vice versa, the presence of a negative weight for a feature indicates that a presence of a particular feature for a user is unlikely, and this could well characterize him.

### C. The matching score

At this point, we may want to match the probe biometrical trait of the user $j$, represented by his positive testing images $X_{\mathrm{te}}^{(j)}$ with the gallery biometrical trait of the user $i$, represented by his positive training images $X_{\mathrm{tr}}^{(i)}$.

Intuitively, a single image does not contain every facet of the visual aesthetics sense of a person; the idea is to consider a *set* of testing images, and guess if the set contains enough information to catch the preferences of the user, allowing to identify him among all the others. Given a template $\mathbf{w}^{(i)}$ of the user $i$, the matching score is aimed at measuring how likely the set $X_{\mathrm{te}}^{(j)}$ of the user $j$ contains images which are in accord with those favorites by the user $i$. In order to determine it, we compute for every image $\mathbf{x}_n^{(j)} \in X_{\mathrm{te}}^{(j)}$ the regression score $\beta_n^{(i,j)}$, as described by eq. 1:

$$\beta_n^{(i,j)} = \mathbf{w}^{(i)T}\ \mathbf{x}_n^{(j)} \tag{4}$$

Then, the final matching score for the whole set (the biometrical trait) is determined as the averaged regression scores of the images belonging to it, i.e.:

$$\beta^{(i,j)} = \frac{1}{N_{\mathrm{te}}} \sum_{n=1}^{N_{\mathrm{te}}} \beta_n^{(i,j)} \tag{5}$$

## III. Experiments

In this section the experimental evaluation is proposed. In particular, we first present the dataset, followed by authentication and recognition results. Finally some interpretability issues are reported.

### A. Data collection

To test our approach, we consider a real dataset of 40000 images, belonging to 200 users chosen at random from the Flickr website. For each user, we retained the first 200 favorites[3]. Please note that the process of adding favorites is a continuous time process, which can last for months. In particular, in our dataset, the minimum amount of time elapsed from the oldest and the newest favorite is 23 weeks (the maximum is 441 weeks) – this ensuring reliable multisession acquisitions. For all the images of the dataset we computed the image signature. In order to guarantee robust testing, we randomly split the images of each user into two parts, one used to build the gallery biometrical trait $X_{\mathrm{tr}}$, and, consequently, its template and one used to build the probe trait $X_{\mathrm{te}}$, needed for testing the algorithm. Since the value ranges are very heterogeneous, each feature is normalized across all training images to have zero mean and unit standard deviation (note that the testing set is normalized with the constants calculated on the training set). In all experiments, the parameter $t$ of the Lasso has been determined by crossvalidation.

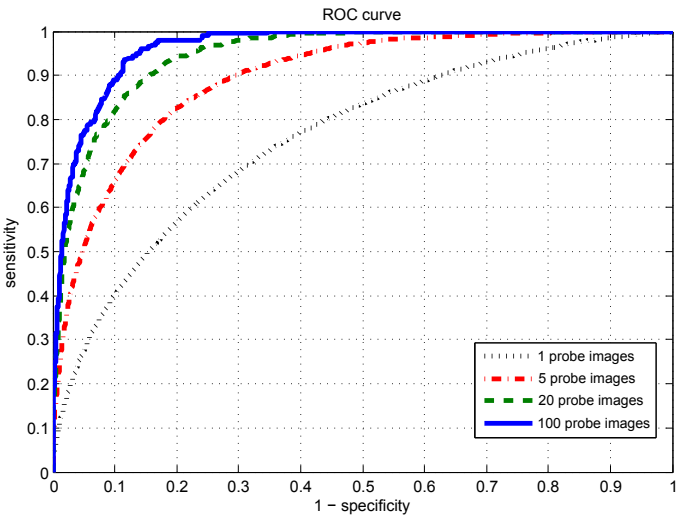
### B. Authentication results

In this section the system is tested in an authentication scenario: a ROC curve is computed for every user $j$, where:

- client images are taken from the probe set of the user $j$
- impostor images are taken from all the other probe sets

In particular, different kinds of client/impostor signatures may be built, depending on the number of images we take into account: in detail, the smallest signature is formed by a single image; pooling together more pictures gives rise to composite signatures, intuitively carrying more information. Matching a signature composed by more than one image occurs by following what is described in Sec. II-C, i.e., roughly speaking, by averaging the matching scores derived from the set of probe images. Given an "authentication threshold", i.e. a value over which the subject is authenticated, sensitivity (true positive

[3]The dataset is available upon request at http://profs.sci.univr.it/~cristanm/projects/perpre.html




| | | | | |
|---|---|---|---|---|
| AUC | 0.76 | 0.89 | 0.95 | 0.96 |
| EER | 0.69 | 0.81 | 0.87 | 0.90 |

Fig. 2. ROC curves for the user authentication, varying the number of images per signature. Each ROC curve has been obtained by averaging over all the ROC curves for each user.

rate) and specificity (true negative rate) can be computed. By varying this threshold the ROC curve is finally obtained.

In Fig. 2 the authentication ROC curves are portrayed; in addition, we reported also the area under the curve (AUC) and the averaged equal error rate (EER), namely, the error when sensitivity and 1-specificity have an equal value. Typically, these values represent a compact and meaningful way to summarize the ROC curve.

As expected, augmenting the images per signatures increments the performance. This confirms the suitability of using this trait as a biometrical trait, even if, as all behavioral biometrics, with a not so outstanding performance.

### C. Recognition results

In this section the recognition capability of the proposed biometrical trait is investigated. In particular, given a probe image or a set of probe images, we want to guess the gallery user who tagged them. To do that, we compute the matching score of the probe image (or set) using all the templates $\{\mathbf{w}^{(i)}\}$. Hopefully, the gallery user with highest score is the one who originally faved the photo (or group of photos).

In order to evaluate the recognition rate, we built a CMC curve [52], a common performance measure in the field of person recognition/re-identification [53]: given a probe set of images coming from a single user and the matching score previously defined, the curve tells the rate at which the correct user is found within the first $k$ matches, with all possible $k$ spanned on the x-axis. Fig. 3 shows various CMC curves for our dataset, where the curves have been obtained by averaging the CMC curves of 20 different experiments with different gallery/probe splits.

On the left, we reported four different CMCs, varying the parameter $N_{\text{te}}$, which tells how many images are aggregated to form a single probe object, while keeping the number of gallery images fixed to 100.

| Protocol | **rank 1** | **rank 5** | **rank 20** | **rank 100** |
|---|---|---|---|---|
| 1 probe image | 0.063 | 0.188 | 0.408 | 0.829 |
| 5 probe images | 0.143 | 0.399 | 0.688 | 0.966 |
| 20 probe images | 0.254 | 0.629 | 0.883 | 0.998 |
| 100 probe images | 0.359 | 0.796 | 0.970 | 0.999 |
| 5 gallery images | 0.076 | 0.236 | 0.496 | 0.889 |
| 10 gallery images | 0.113 | 0.322 | 0.628 | 0.948 |
| 20 gallery images | 0.152 | 0.443 | 0.743 | 0.971 |
| 50 gallery images | 0.225 | 0.578 | 0.838 | 0.995 |

TABLE II
CMC VALUES FOR DIFFERENT RANKS. VALUES REPRESENT THE PROBABILITY OF HAVING THE CORRECT MATCH WITHIN THE FIRST 1-5-20-100 SIGNATURES, CONSIDERING DIFFERENT NUMBERS OF PROBE (FIRST 4 ROWS OF THE TABLE) AND GALLERY IMAGES (LAST 4 ROWS OF THE TABLE).

From the figure it is evident that performing the task of identifying correctly a user with a single image (black dotted line) is very difficult. However, as soon as the number of probe images grouped together increase a little, a consistent improvement can be noted. This is in line with our hypothesis: we are aggregating information from heterogeneous images, each one characterizing only a small portion of the user subjective tastes.

On the right, we assessed the importance of the gallery set size $N_{\text{tr}}$ by keeping the probe parameter $N_{\text{te}}$ fixed to 20. As expected, by lowering the number of gallery elements, it is more difficult to learn the users' preferences and their aesthetic sense uniqueness. For both figures, the normalized Area Under the Curve (nAUC) has been reported in the legend.

As a further comment, it is also worth noting that, even if at CMC rank 1 we achieve in the worst case a 6.3% rate of correct identification, this is higher than the probability of recognizing the user by mere chance (which amounts to 0.5%).

A final interesting question can be made: how is our signature when compared to other biometrical cues? Having clear in mind that realizing a proper and exhaustive comparison is not so trivial, we would like to provide here some intuitions. We focus on the field of people re-identification, where the signature of a user is composed by set of his full-body images (i.e., the appearance). In particular, we take into account the experiment done on the CAVIAR4REID dataset in [53], where multiple methods of re-identification have been tested on small images of people (averagely, around $50 \times 120$ pixel). In order to create a fair comparison with the CAVIAR4REID re-identification experiment, we randomly select the same number of users (72) and we used 5 images for the gallery, 5 for the probe, repeating the experiments 10 times. In our case, we obtain an nAUC of 74.8%, whereas with the classical re-identification approaches the n-AUC (reported in the paper) is 78.5%. This result is quite intriguing, as it states that having images chosen and marked by an user as his favorites is not too far from actually looking at that subject directly. This witnesses the potentiality of our biometrical strategy.

### D. Feature analysis

This section is aimed at providing a qualitative evaluation of the proposed approach, showing that the regression score

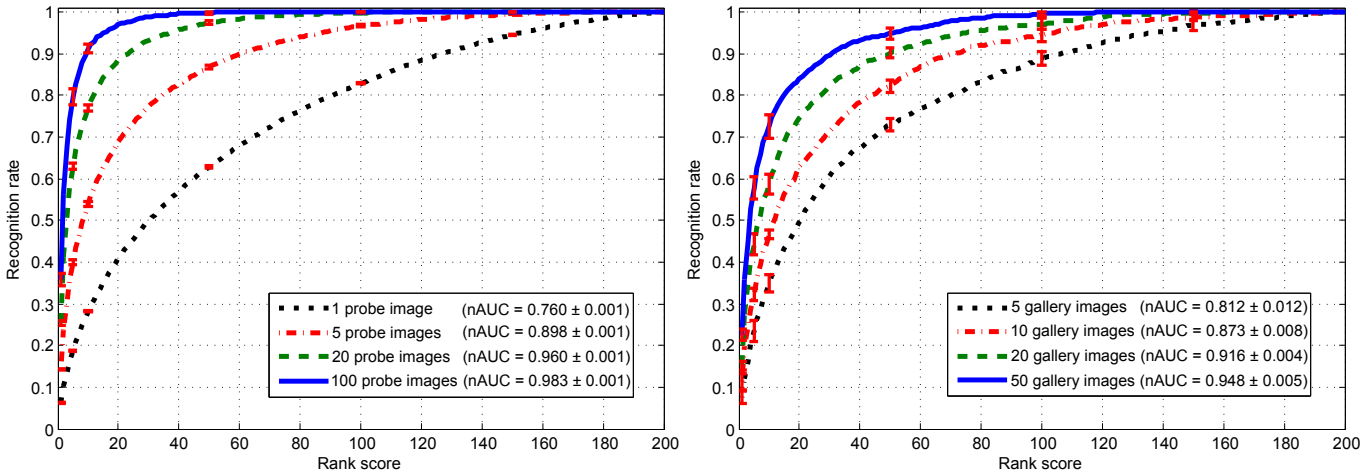



Fig. 3. CMC curves for our dataset: the curves have been obtained by averaging the CMC curves of 20 different experiments with different gallery/probe splits. On the left: for each curve, we varied the number of probe images to be considered as a single "set", while keeping the number of gallery images fixed to 100. On the right: for each curve, we varied the number of gallery images used to train Lasso, while keeping the number of probe images to 20. Since we performed 20 random splits of gallery/probe, we report also the standard deviation of results. Table II reports more in detail the values of both curves at rank 1-5-20-100, in order to provide a better quantitative idea of the probability of having the correct match within the first 1-5-20-100 signatures.

$\beta$ provides a valid measure of the preferences of a user, while the weight coefficients in the vector $\mathbf{w}$ provide an interpretable description for his visual aesthetic sense. In the first experiment, given the gallery user $i$, we considered all the probe images of all the users $\{j\}$, and we sorted them according to their regression score $\beta^{(i,j)}$. The higher the score of an image, the higher the probability that the user may have actually faved that image. Fig. 4 gives an excerpt of the results; each column corresponds to a different Flickr user $i$; given the template of that user, the first 10 rows are the favorite gallery photos which exhibit high regression scores, ranked in descending order from the highest one. In other words, these 10 images are the ones which better represent the user $i$, as modeled by the related template. The second 10 images are taken from the probe images of all the users (not only from user $i$), which exhibit the 10 highest regression scores (again, w.r.t the template of the user $i$), with a blue frame indicating actually those images which belong to that user.

The figure reveals some interesting information: although the highest test image for the template of the user is not on his favorites set, it can have some visual appeals reflected on some of the images on his gallery set (see for example the black and white faces and scenes in the first column, or the airplanes in the second). It seems that a sort of "internal coherence" starts to show up.

We then looked into the weight coefficients for some users after learning the sparse regression model. For 2 random users, we reported the vector $\mathbf{w}$ in Fig. 5, on the right of their gallery and predicted preferred images. For visualization purposes, we labeled the most prominent features (i.e. the ones with highest – in absolute value – weight value):

For user 41, the rule of thirds (i.e., its computational aesthetics version) plays an important role, and actually most of his images report an object in the central rectangle of the image. This is visible also in the probe images, selected by regression among the probe images of all the users; all the images appear with high luminosity, and the same happens in the probe images of that user. Note also that there are few regions in the gallery/probe images, this being reflected by the corresponding negative weight. For user 182, faces, the white color, hue homogeneity and edge/textural properties are important, and this is visible since many black/white images of many people (many edges/textures) characterize his favorite set of shots. Similarly, probe images which conform with the classifier of that user report faces and edgy pictures, some of them with few colors (high hue homogeneity), with a strong presence of white. The negative weight on aspect ratio indicates that images composing the favorite set of user 182 are more "rectangular" than the preferred images of others.

## IV. Conclusions

The key idea of this article is that the cognitive mechanisms that regulate the appreciation of an image are personal and unique, and that their distillation can provide an interesting soft-biometrical trait. To this aim, we used a typical learning approach, considering images tagged as "favorite" by a certain person as training data, incorporating the expression of her aesthetic preferences. In the experiments we validated this intuition with a consistent amount of pictures taken from Flickr, showing thus that personal aesthetic traits may be collected and managed easily from the social web. To the best of our knowledge, such a perspective has never been adopted in a forensic technology context before. The probable reason is that multimedia data became an interaction channel only recently, when the diffusion of appropriate technologies for data production (cameras, smartphones, tablets, etc.) and consumption (social media, digital libraries, etc.) made it possible to exchange multimedia data as easily as we previously exchanged written material (letters, messages, etc.) [54]. Our proposal could be interesting for several aims: apart from the design of a classical biometric system, where personal aesthetics may support other stronger biometrical cues, other applications can be thought of: for example, exploiting eye tracking devices, the spatio-temporal patterns with

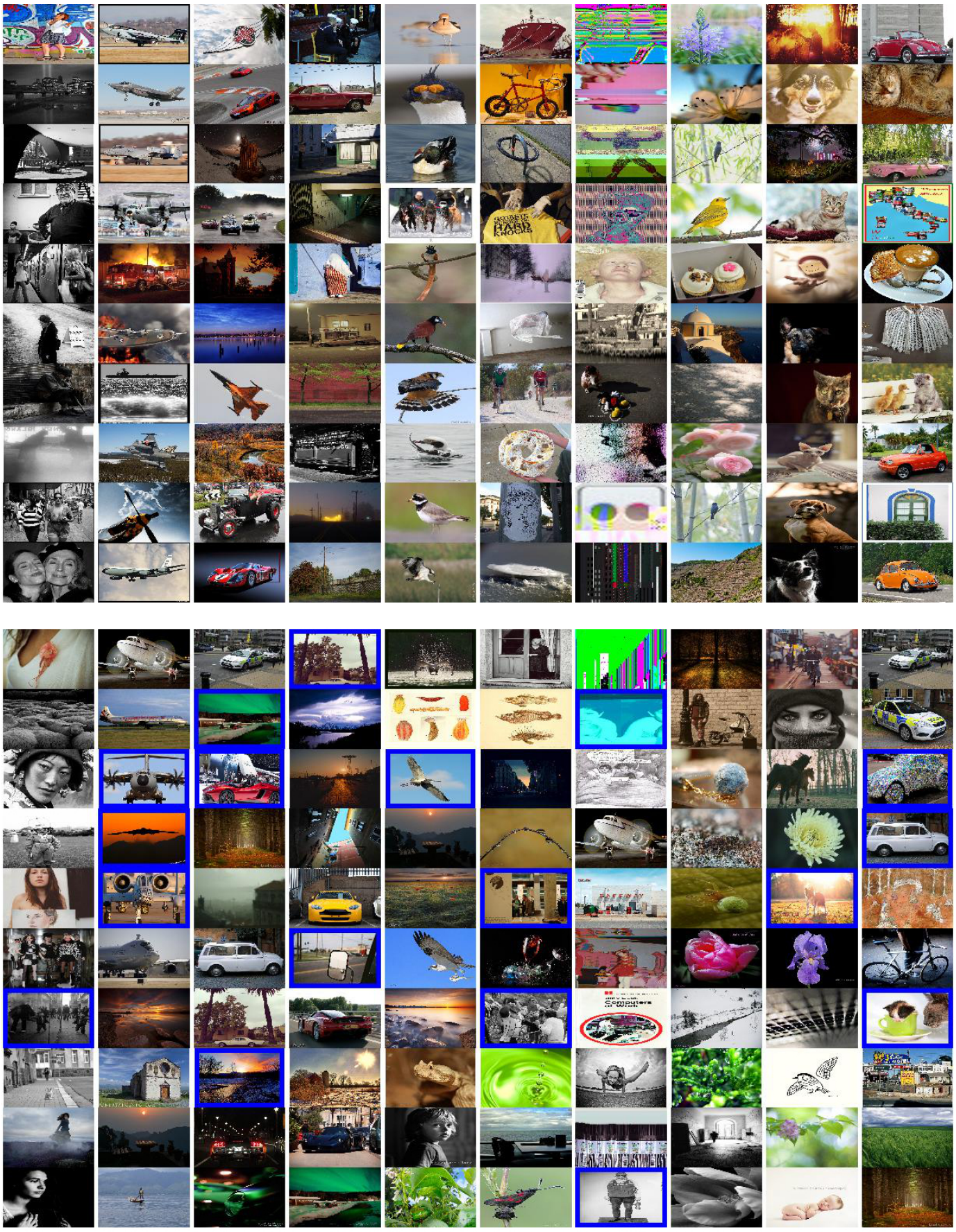


Fig. 4. Gallery and recognized probe images for different users. Each column is a user, and the first 10 images come from his gallery set. In the half-bottom part, we show the first 10 probe images for that user, ranked on the basis of their regression score (the first being the one with highest score). In blue, correct matches are highlighted. A "coherence" between gallery and probe images can be seen.

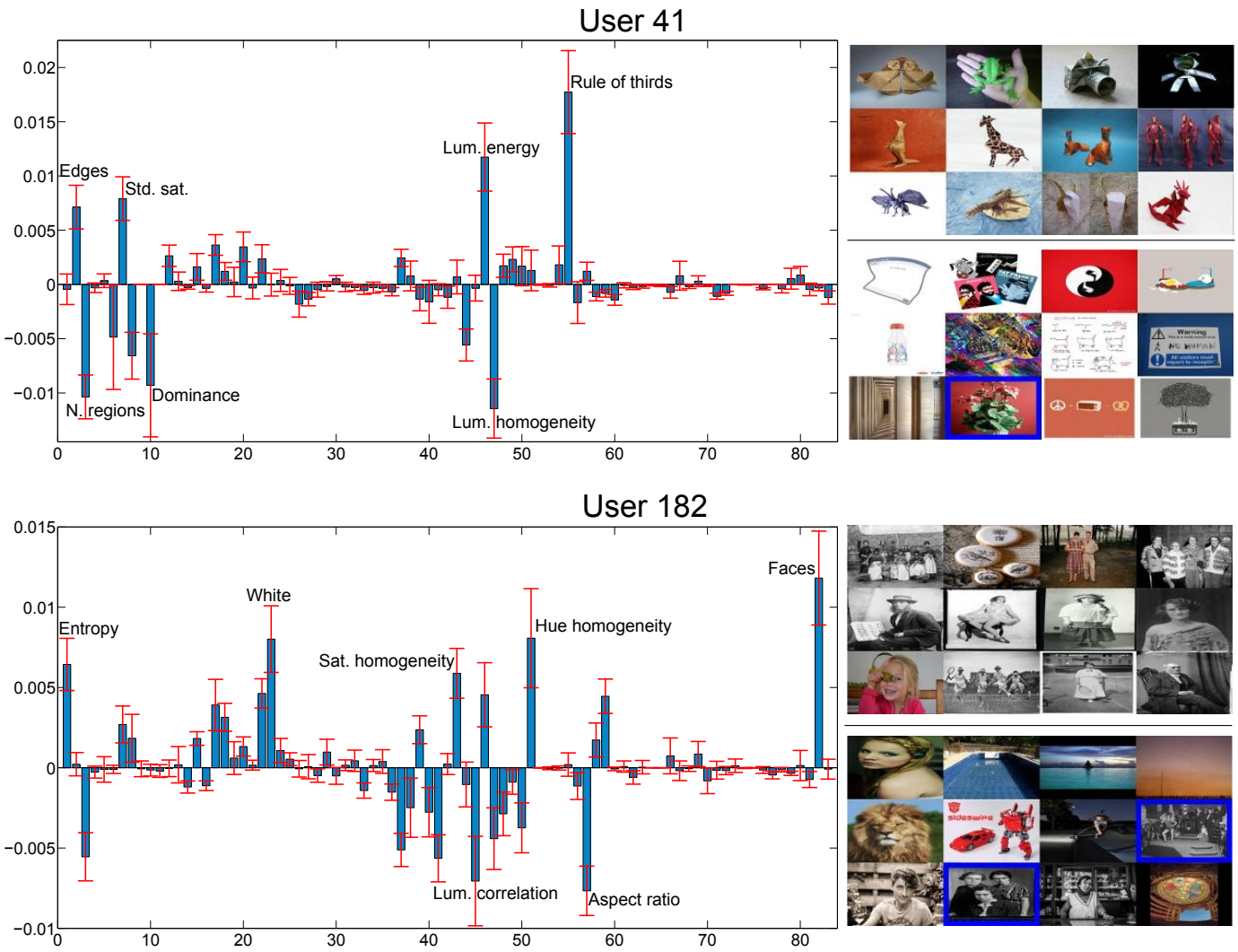



Fig. 5. Most prominent features for 2 users taken from the dataset. On the left, for each user, a bar plot of each feature's importance is shown. The height of each bar represents the value of the corresponding weight, and the standard deviation along the 20 different experiments on different partitions is also visualized. On the right, gallery and probe elements are shown, in the same fashion of Fig. 4.

which preferred images are explored may enrich and reinforce the biometrical traits. Even more pioneering, analyzing the subjective preferences of a person may unveil the interplay between personality traits and image features [55], which can bring in ethical aspects ignored so far. As future work, we will analyze different user populations, exploring how socially interrelated users may exhibit similar aesthetical signatures: for example, the judgement of attractiveness of a face has been shown to be correlated among people connected by tight personal relations, as siblings, spouses, close friends [21]; this may have a negative impact on our biometrical strategy, augmenting the probabilities of breaking our biometric system. Technically, we will take into account the rule that each image may have in the definition of the aesthetical preferences of a user: not all the images will intuitively have the same importance, and this could be modeled by Multiple Instance Learning techniques, recently employed for person recognition tasks [56]. We will also design prototypes of authentication and recognition interfaces, so that user usability studies can be performed, toward a real deployment of this new biometric strategy.

## References


[1] A. Jain, P. Flynn, and A. Ross, *Handbook of Biometrics*. Springer, 2008.

[2] M. Tistarelli, S. Li, and R. Chellappa, *Handbook of Remote Biometrics for Surveillance and Security*. Springer, 2009.

[3] S. Li and A. Jain, *Handbook of Face Recognition*. Springer, 2005.

[4] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar, *Handbook of Fingerprint recognition*, 2nd ed. Springer, 2009.

[5] M. Burge and K. Bowyer, *Handbook of Iris Recognition*. Springer, 2013.

[6] A. Abaza, A. Ross, C. Hebert, M. Harrison, and M. Nixon, "A survey on ear biometrics," *ACM Comput. Surv.*, vol. 45, no. 2, pp. 22:1–22:35, 2013.

[7] H. Chen and A. Jain, "Dental biometrics: Alignment and matching of dental radiographs," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1319–1326, 2005.

[8] A. Riera, M. Soria-Frisch, C. Caparrini, C. Grau, and G. Ruffini, "Unobtrusive biometric system based on electroencephalogram analysis," *EURASIP Journal on Advances in Signal Processing*, vol. 1, pp. 18–25, 2008.

[9] R. Yampolskiy and V. Govindaraju, "Behavioural biometrics: a survey and classification," *Int. J. Biometrics*, vol. 1, no. 1, pp. 81–113, 2008.

[10] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "A review of vision-based gait recognition methods for human identification," in *Int. Conf. on Digital Image Computing: Techniques and Applications (DICTA)*, 2010, pp. 320–327.

[11] M. Nixon, T. Tan, and R. Chellappa, *Human Identification Based on Gait*. Springer-Verlag, New York Inc., 2006.

[12] Y. Zhu, T. Tan, and Y. Wang, "Biometric personal identification based on handwriting," in *Int. Conf. on Pattern Recognition*, vol. 2, 2000, p. 2797.

[13] G. Bailador, C. Sanchez-Avila, J. Guerra-Casanova, and A. de Santos Sierra, "Analysis of pattern recognition techniques for in-air signature biometrics," *Pattern Recognition*, vol. 44, no. 1011, pp. 2468 – 2478, 2011.

[14] M. Pusara and C. Brodley, "User re-authentication via mouse movements," in *ACM workshop on Visualization and data mining for computer security*. ACM, 2004, pp. 1–8.

[15] M. Rybnik, M. Tabedzki, and K. Saeed, "A keystroke dynamics based system for user identification," in *Int. Conf. on Computer Information Systems and Industrial Management Applications*, 2008, p. 225230.

[16] L. Olejnik, C. Castelluccia, and A. Janc, "Why johnny can't browse in peace: On the uniqueness of web browsing history patterns," in *Proc. of W. on Hot Topics in Privacy Enhancing Technologies*, 2012.

[17] G. Roffo, C. Segalin, A. Vinciarelli, V. Murino, and M. Cristani, "Reading between the turns: Statistical modeling for identity recognition and verification in chats," in *IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, 2013.

[18] H. Leder, B. Belke, A. Oeberst, and D. Augustin, "A model of aesthetic appreciation and aesthetic judgments," *British Journal of Psychology*, vol. 95, no. 4, pp. 489–508, 2004.

[19] C. Martindale, K. Moore, and J. Borkum, "Aesthetic preference: Anomalous findings for berlyne's psychobiological theory," *American Journal of Psychology*, vol. 103, no. 1, pp. 53–80, 1990.

[20] U. Eco, *History of beauty*. Rizzoli Intl Pubns, 2004.

[21] P. Bronstad and R. Russell, "Beauty is in the "we" of the beholder: greater agreement on facial attractiveness among close relations." *Perception*, vol. 36, no. 11, pp. 1674–1681, 2007.

[22] S. Bhattacharya, R. Sukthankar, and M. Shah, "A framework for photo-quality assessment and enhancement based on visual aesthetics," in *International conference on Multimedia*. New York, NY, USA: ACM, 2010, pp. 271–280.

[23] E. A. Vessel and N. Rubin, "Beauty and the beholder: Highly individual taste for abstract, but not real-world images," *Journal of vision*, vol. 10, no. 2, 2010.

[24] I. McManusU, A. L. Jones, and J. Cottrell, "The aesthetics of colour," *Perception*, vol. 10, pp. 651–666, 1981.

[25] B. Adams, "Where does computational media aesthetics fit?" *IEEE Multimedia*, vol. 10, pp. 18–27, 2003.

[26] I. McManus, R. Cook, and A. Hunt, "Beyond the golden section and normative aesthetics: Why do individuals differ so much in their aesthetic preferences for rectangles?" *Psychology of Aesthetics, Creativity, and the Arts*, vol. 4, no. 2, p. 113, 2010.

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

[28] R. Datta, D. Joshi, J. Li, and J. Wang, "Studying aesthetics in photographic images using a computational approach," in *European Conference on Computer Vision*. Springer Berlin / Heidelberg, 2006, vol. 3953, pp. 288–301.

[29] H.-H. Su, T.-W. Chen, C.-C. Kao, W. H. Hsu, and S.-Y. Chien, "Preference-aware view recommendation system for scenic photos based on bag-of-aesthetics-preserving features," *IEEE Trans. on Multimedia*, vol. 14, no. 3-2, pp. 833–843, 2012.

[30] Y. Ke, X. Tang, and F. Jing, "The design of high-level features for photo quality assessment," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2006, pp. 419–426.

[31] Y. Luo and X. Tang, "Photo and video quality evaluation: Focusing on the subject," in *European Conference on Computer Vision*. Springer-Verlag, 2008, pp. 386–399.

[32] I. Biederman and E. Vessel, "Perceptual pleasure and the brain," *American Scientist*, vol. 94, no. 3, pp. 1–8, 2006.

[33] A. Bozzon, M. Brambilla, and S. Ceri, "Answering search queries with crowdsearcher," in *WWW*, 2012, pp. 1009–1018.

[34] P. Lovato, A. Perina, N. Sebe, O. Zandonà, A. Montagnini, M. Bicego, and M. Cristani, "Tell me what you like and I'll tell you what you are: discriminating visual preferences on Flickr data," in *Asian Conference on Computer Vision*, 2012.

[35] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," in *International Conference on Multimedia*. ACM, 2010, pp. 83–92.

[36] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, "Discriminatively trained deformable part models, release 4," http://www.cs.brown.edu/~pff/latent-release4/, 2010.

[37] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2001, pp. 511–518.

[38] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.

[39] P. Valdez and A. Mehrabian, "Effects of color on emotions." *J. Experimental Psychology Gen.*, vol. 123, no. 4, pp. 394–409, Dec. 1994.

[40] K. Mardia and P. Jupp, *Directional Statistics*. Wiley, 2009.

[41] H. Tamura, S. Mori, and T. Yamawaki, "Texture features corresponding to visual perception," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 8, no. 6, 1978.

[42] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603 – 619, 2002.

[43] C. Georgescu, "Synergism in low level vision," in *International Conference on Pattern Recognition*, 2002, pp. 150–155.

[44] J. van de Weijer, C. Schmid, and J. Verbeek, "Learning color names from real-world images," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[45] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, 1st ed. Addison-Wesley Longman Publishing Co., Inc., 1992.

[46] P. Isola, J. Xiao, A. Torralba, and A. Oliva, "What makes an image memorable?" in *IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '11. IEEE Computer Society, 2011, pp. 145–152.

[47] W. Curran, T. Moore, T. Kulesza, W. Wong, S. Todorovic, S. Stumpf, R. White, and M. M. Burnett, "Towards recognizing "cool": can end users help computer vision recognize subjective attributes of objects in images?" in *ACM Int. Conf. on Intelligent User Interfaces*, 2012, pp. 285 –288.

[48] R. B. Girshick, P. F. Felzenszwalb, and D. McAllester, "Discriminatively trained deformable part models, release 5," http://people.cs.uchicago.edu/ rbg/latent-release5/, 2010.

[49] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part based models," *IEEE Trans. on Pattern Analysis and Machine Intelligence"*, vol. 32, no. 9, pp. 1627–1645, 2010.

[50] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. I–511–I–518, 2001.

[51] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.

[52] H. Moon and P. Phillips, "Computational and performance aspects of pca-based face-recognition algorithms," *Perception*, vol. 30, pp. 303 – 321, 2001.

[53] L. Bazzani, M. Cristani, and V. Murino, "Symmetry-driven accumulation of local features for human characterization and re-identification," *Computer Vision and Image Understanding*, vol. 117, no. 2, pp. 130–144, Feb. 2013.

[54] A. Burdick, J. Drucker, P. Lunenfeld, T. Presner, and J. Schnapp, *Digital_Humanities*, M. Press, Ed., 2012.

[55] M. Cristani, A. Vinciarelli, C. Segalin, and A. Perina, "Unveiling the multimedia unconscious: Implicit cognitive processes and multimedia content analysis," in *ACM Multimedia Conference (ACMMM)*, 2013, brand New Idea paper.

[56] R. Satta, G. Fumera, F. Roli, M. Cristani, and V. Murino, "A multiple component matching framework for person re-identification," in *16th international conference on Image analysis and processing (ICIAP 2011)*, ser. ICIAP'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 140–149.