Person Recognition using a bag of facial soft biometrics (BoFSB)

Antitza Dantcheva, Jean-Luc Dugelay and Petros Elia

Multimedia Communications Department,
EURECOM, Sophia Antipolis, France
{Antitza.Dantcheva, Jean-Luc.Dugelay, Petros.Elia}@eurecom.fr

***Abstract*— This work introduces the novel idea of using a bag of facial soft biometrics for person verification and identification. The novel tool inherits the non-intrusiveness and computational efficiency of soft biometrics, which allow for fast and enrolment-free biometric analysis, even in the absence of consent and cooperation of the surveillance subject. In conjunction with the proposed system design and detection algorithms, we also proceed to shed some light on the statistical properties of different parameters that are pertinent to the proposed system, as well as provide insight on general design aspects in soft-biometric systems, and different aspects regarding efficient resource allocation.**

I. INTRODUCTION

A. Soft-biometrics: origins and utility

Soft-biometric considerations were first analytically introduced by Bertillon (see for example [1]) who brought to the fore the idea of using biometric, morphological and anthropometric determinations for person identification. A great majority of Bertillon's early descriptors/traits currently fall under the category of soft biometrics, as defined by Jain et al. [3] to be a set of characteristics that provide some biometric information, but are not able to individually authenticate the person, mainly due to lack of distinctiveness and permanence.

Currently state–of–the–art systems on intrusion detection and security mechanism, include by default at least one biometric trait. Incorporating soft biometrics can increase the system reliability, and can impart substantial advantages: soft biometric features reveal biometrical information of individuals, they can be partly derived from the main biometrics identifier, they are easily captured, and their acquisition does not require enrolment since training can be performed in advance on individuals out of the specific authentication group. As further noted in [2], soft biometrics are not expensive to compute, can be sensed at a distance in a crowded environment, do not require the cooperation of the surveillance subjects, and are especially useful in narrowing down the search within a group of candidate individuals for face recognition. More recent work on soft biometrics has been performed predominantly with the aim of pre-processing for face recognition. Recent work on using soft-biometrics for verification can be found in [11] which though, unlike the current work, focuses instead on *body* soft biometrics.

B. Novel bag of facial soft biometrics, and related applications

Motivated by the above plethora of utilities, the present work develops a tool for detection of facial soft biometric traits that emphasizes on the most obvious facial identifiers, primarily mentioned by humans, when portraying an unknown individual. The constructed tool is streamlined to achieve reliability of authentication at reduced complexity, and hence focuses on a specific set of simple yet robust soft-biometric traits, including hair color, eye color and skin color, as well as the existence of beard, moustache and glasses.

In conjunction with the proposed system, the current work also presents some aspects for soft-biometric system design, as well as statistical characterization of relevant parameters, and an analysis of capabilities and limitations of general soft-biometric systems.

The constructed tool has several diverse applications, including:

- Expediting face recognition by search pruning
- Re-identifying a described criminal
- Searching surveillance videos for suspects

Other applications relate to the ability to match people based on their biometric-trait preferences, acquiring statistical properties of biometric identifiers of groups, avatar modelling based on the instantaneous facial characteristics (glasses, beard or different hair color), statistical sampling of audiences, and many others.

II. GENERAL SETTING OF SOFT-BIOMETRIC AUTHENTICATION: PERTINENT MEASURES AND PARAMETERS

A. Main parameters: authentication group, traits, trait-instances, and categories

The setting of interest corresponds to the general scenario where, out of a large population, an *authentication group* is randomly extracted as a random set of N people, out of which one person is picked for authentication (and differentiation from all the other members of the authentication group). We note that this general scenario is compliant with both, the case of person verification as well as of identification. A general soft-biometric system employs detection that relates to λ *soft-biometric traits* (hair color, skin color, etc), where each trait i, i=1, 2,… , λ, is subdivided into μ_i *trait-instances*, i.e., each trait i can take one of μ_i values. We henceforth denote as *category* to be any λ-tuple of different trait-instances, and we let $\Phi=\{\varphi_i\}_{i=1}^{\rho}$ define a set of all ρ categories, i.e., the set of all

ρ combinations of soft-biometric trait-instances. The number of categories ρ, that the system is endowed with, is given by

$$\rho = \prod_{i=1}^{\lambda} \mu_i. \quad (1)$$

We slightly abuse notation and henceforth say that *a subject belongs in category φ* if his or her trait-instances are the λ-tuple corresponding to category φ. We here note that to have conclusive authentication of a subject, and subsequent differentiation from the other subjects of the authentication group, it must be the case that the subject does not belong in the same category as other members of the authentication group.

Given a specific authentication group, the maximum-likelihood (ML) optimizing rule for detecting the most probable category $\hat{\varphi} \in \Phi$ in which a chosen subject belongs, is given by:

$$\hat{\varphi} = arg \max_{\varphi \epsilon \Phi} P(\varphi)\, P(y/\varphi), \quad (2)$$

where y is the observation vector, P(φ) is the pdf of the set of categories over the given population (note $\sum_{\varphi=1}^{\rho} P(\varphi) = 1$), and P(y/φ) the probability that y is observed, given that the subject belongs in category φ.

B. Design aspects in soft-biometric systems

In designing a soft-biometric system, the overall choice of the traits and trait-instances, must take into consideration aspects as traditional limitations on estimation reliability, which is commonly a function of the sensor resolution, and of the capabilities of the image-processing part of detection. In addition to this traditional aspect, new concerns come into the picture when designing a soft-biometric system as of the size and statistics of the authentication group (such as the possible similarities that might exist between different subjects), as well as the statistical relationship between the authentication group and Φ. The interrelated nature of the above aspects brings to the fore different tradeoffs. Such tradeoffs include for example the fact that an increasing μ_i, and thus also an increasing ρ, generally introduce a reduction in the reliability of detection, but can potentially result in a welcomed increase in the maximum authentication group size (N) that the system can accommodate for.

It then becomes apparent that designing and analysing soft-biometric systems requires a deviation from traditional design and analysis of classical multi-biometric systems, towards considering the role of the above parameters, and their effect on the tradeoffs and the overall system performance. This approach motivates the proposed soft-biometric system design described in Section III, as well as the subsequent system analysis of Section IV which also includes simulation evaluation of the proposed system in the interference limited setting of very high sensor resolution.

III. The Proposed Soft-Biometric System

In accordance with the above design aspects, and in an effort to find a good balance between authentication-reliability and complexity, we here propose a soft-biometric system that focuses on *simple and robust* detection from a bounded set of traits and their trait-instances. In what follows, we will describe these basic elements, as well as the employed detection algorithms.

A. Chosen features of the proposed soft-biometric system

In the presented bag of facial soft biometric traits for human authentication, we allocate λ = 6 traits, which we choose and label as:

1. skin color
2. hair color
3. eye color
4. presence of beard
5. presence of moustache
6. presence of glasses.

In this setting we clearly assign $\mu_4 = \mu_5 = \mu_6 = 2$, corresponding to the binary nature of traits i = 4, 5, 6. On the other hand, the first three traits are of a continuous character (see Table I) and had to be categorized in consideration to the tradeoff between reliability of detection and trait importance. Towards this we chose to subdivide trait 1 (skin color) into $\mu_1 = 3$ instances and label them (following a recommendation provided by the ethical partner of an ongoing EU project, ActiBio [16] to avoid any assumptions about race or ethnicity based on skin color) as:

- { skin color type 1, skin color type 2, skin color 3 } using numbers that increase from light to dark,

to subdivide trait 2 (hair color) into $\mu_2 = 8$ instances

- { light-blond, dark-blond, brown-, black-, red-, grey-, white-haired, and bald }

and to subdivide trait 3 (eye color) into $\mu_3 = 6$ instances:

- {blue-, green-, brown-, grey-, mixed-, black-eyed }.

As a result, the proposed system is endowed with the ability to detect

$$\rho = \prod_{i=1}^{6} \mu_i = 1152 \quad (3)$$

distinct categories.

For the sake of clarification, we note two simple examples of such categories in Φ:

- "skin type 1, brown hair, blue eyes, no beard, no moustache, no glasses" ϵ Φ
- "skin type 3, black hair, black eyes, beard present, moustache present, glasses present" ϵ Φ.

Having described the basic parameters of the system, we proceed to specify basic aspects of the detection algorithms that were used for trait-instance identification.

B. Detection algorithms

The basic detector consisted of an automatic frontal face and facial features detector, which was partially drawn and modified from the algorithms in [7]. Implementation of the different detection algorithms (see Table I for an overview) was performed using OpenCV [10].

TABLE I
FACIAL SOFT BIOMETRICS ALGORITHMS OVER VIEW

Soft biometric trait	**Algorithm**	**Database**	**Nature of value**
Skin color	Deduced from [8]	[6]	Continuous
Hair color	Deduced from [9]	[6]	Continuous

Eye color	Own developed	[5]	Continuous
Beard	Own developed	[6]	Binary
Moustache	Own developed	[6]	Binary
Eye glasses	Deduced from [4]	[6]	Binary

Pre-detection aspects: Before describing some basic aspects of the implemented trait detection algorithms, we note a few pertinent issues that accompany detection. Regarding coordinate determination, we note that typical eye, skin and hair color detectors require knowledge of the eye coordinates, and similarly hair color detection requires knowledge of the coordinates for the upper head region. The precise computation and extraction of the characteristic regions of interest (ROI) (see Figure 1) for the eyes, mouth, nose and upper face coordinates, are essential for the subsequent detection. For higher accuracy, only in the training step, all coordinates were manually annotated. The considered ROIs for the selected soft biometric traits are illustrated in Figure 1.

Identification of the ROI was generally followed by acquisition of the Hue, Saturation and Value (HSV) values. We note that the HSV color-space was chosen for being robust to illumination changes, as well as for the fact that it allows for a high degree of independence between the H, S, and V parameters, which renders the system capable to better handle light changes or shadows. Regarding outlier filtering, we used a simple threshold on the HSV values, based on the color standard-deviation σ. This was followed by HSV normalization. Regarding the statistical modelling, the probability density functions of skin, eye and hair color were computed using 3-component Gaussian mixture models whose parameters were estimated using the EM algorithm. Posterior probabilities over the observed HSV vectors for all trained trait instances were computed, followed by a majority vote decision on the detected trait instance.

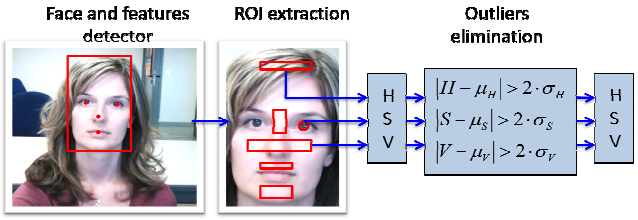

Figure 1. ROI for the bag of facial soft biometrics. Outlier filtering was a function of the standard deviation σ and the mean μ for each of the H,S and V parameters.

1) *Eye Color Detection:* In this setting, careful and precise consideration of the ROI was particularly important, due to the region's inherently small size. The specific ROIs were retrieved using the circular Hough transformation**,** followed by pupil and reflection extraction, and then by acquisition of the HSV vectors. Regarding the training step, each eye color group was trained using images from the Caltech database [5].

2) *Hair color detection:* The hair color ROI was chosen as a thin bar in the upper head region, as indicated in Figure 1. Training utilized 30 Feret images for each of the hair colors, where the classification was done manually.

3) *Skin color:* Detection of skin color was done in accordance to the eye coordinates which defined the ROI for the skin color detection to be the area underneath the ocular region. Training utilized 33 Feret images per skin color group, which were again annotated manually.

4) *Eye glasses detection:* Towards glasses detection we considered that the areas around the eyes can be searched both for hints of glasses as well as for glass reflections. Challenges related to the fact that glass frames are either occasionally absent, or that they often resemble wrinkles, brows, shades and hair. Further challenge came from the fact that illumination variances hindered the appearance of reflections. These challenges were handled by placing emphasis on a ROI corresponding to the nose part of the glasses. The specific algorithm consisted of eye position detection, grey-level conversion, histogram equalization, extraction of region between the eyes, Laplacian edge detection, and finally line detection.

5) *Beard and Moustache Detection:* In this case, face detection and feature-localization were followed by identification of the ROIs. These ROIs include the chin for the beard, and the area between the mouth and nose for the moustache. The color estimation was followed by outlier extraction and HSV normalization. The presence of beard and/or moustache was based on the Euclidean distance between the processed observation and skin- and hair-color information respectively. The presence of moustache was determined independently.

6) *Algorithmic dependencies:* As it is the case with general optimization problems, identification of algorithmic dependencies endows the system with increased reliability and computational efficiency. Towards this we refer to notable examples of such dependencies, such as that between skin color and glasses where, due to ROI overlapping, the presence of glasses has an impact on the perceived skin color. This information can be utilized and employed by modifying the ROI for skin color detection. Additionally we recall that skin color is employed in the detection of hair, beard and moustache, where furthermore the latter two traits are also contingent upon hair color.

Figure 2 sketches further dependencies of the mentioned facial soft biometric traits. Some of these dependencies were partly exploited in the process of detection.

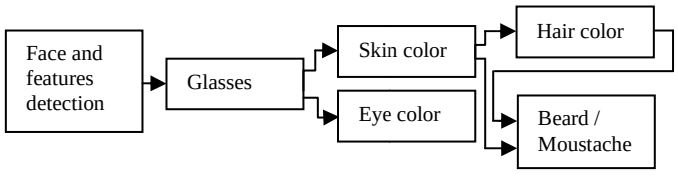

Figure 2. Facial soft biometric traits dependencies

Having described different aspects of the proposed soft-biometric system design, we proceed to shed some light on the statistical properties of the pertinent parameters and to provide some early results on the statistical characterization corresponding to these aspects of the proposed scheme, as well as simulations in the interference limited setting of very high sensor resolution.

IV. STATISTICAL ASPECTS OF THE CONSTRUCTED SCHEME

Relevant parameters, in addition to λ, $\{\mu_i\}$, and ρ, also include the size and statistics of the authentication group (revealing possible similarities between different subjects), as well as the statistical relationship between the authentication group and Φ. In what follows we aim to gain insight on the behaviour of the above, in the specific setting of the proposed soft-biometric design. The following analysis, which is by no means conclusive, focuses on providing insight on parameters such as:

- The spread of the effective categories for a given authentication group, where this spread is used as a measure of the suitability of Φ in authenticating subjects from a certain authentication group.
- The relationship between N, and the corresponding probability of interference as a function of Φ (the probability that two users share the same category and will thus be indistinguishable).
- The probability of interference-induced authentication error, again to be considered as a measure of the system's reliability).

A. *Spread of the category set Φ*

We here consider the case where a soft-biometric system is designed to distinguish among ρ distinct categories, but where the randomly introduced authentication group only occupies a smaller fraction of such categories, and where these categories are themselves substantially correlated. Leaving correlation issues aside for now, we first define the set of *effective categories* Φ_e to be the set of categories that are present (are non empty) in the specific authentication group. A pertinent measure of system diversity and performance then becomes the cardinality $\rho_e = |\Phi_e|$. We note that clearly both Φ_e and ρ_e are random variables, whose realizations may change with each realization of the authentication group.

To gain insight on the above randomness, we consider the case where the authentication groups are each time drawn from general population that is a fixed set of $K = 646$ subjects taken from the Feret database [6], with $\rho = 1152$ categories, corresponding to a pdf $P(\varphi)$ as shown in Figure 3, where this pdf itself corresponds to the traits and trait-instances of the proposed system.

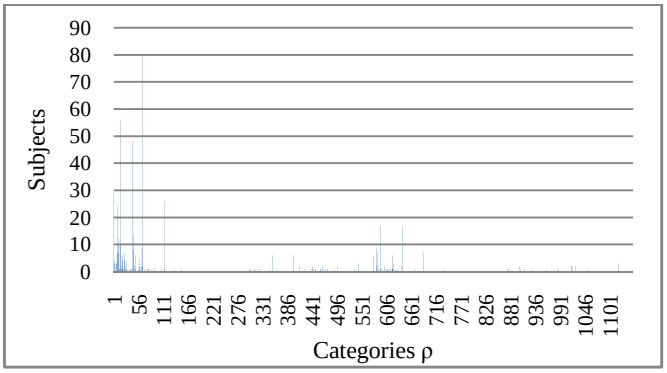

Figure 3. $P(\varphi)$ corresponding to Feret distribution and the proposed system

Given the above, Figure 4 describes the average number of empty categories,

$$\rho - E[\rho_e](N), \tag{4}$$

as a function of N, where the expectation is taken over the different realizations of authentication groups.

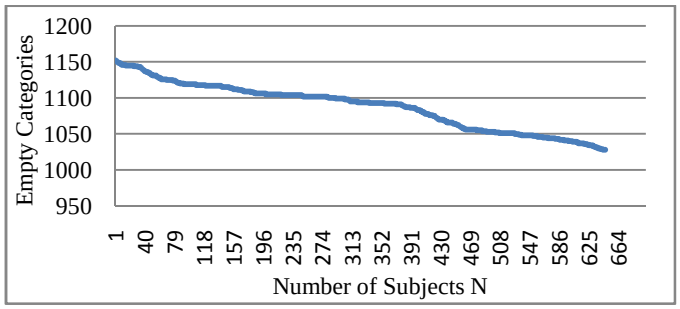

Figure 4. Expected number of empty categories as a function of N (Feret)

It becomes apparent that a natural remedy for increasing $E[\rho_e]$ is to increase the overall ρ, which brings to the fore the natural question as to whether this increase in ρ should be more a result of an increase in the number of traits, or rather more a result of the number of trait-instances. We address this resource allocation problem, under the simplifying assumption of symmetry, where $\mu_i = \mu$, for all $i=1,\ldots,\lambda$. In this symmetric setting, where clearly

$$\rho = \mu^{\lambda} \tag{5}$$

and where ρ increases polynomially with μ and exponentially with λ, a simple comparison of the two derivatives $\frac{d\rho}{d\mu}, \frac{d\rho}{d\lambda}$, identifies the *trait-limited region* of a soft-biometric system to be the region

$$\lambda < \mu \ln \mu \tag{6}$$

in which ρ increases faster with λ than with μ, and where emphasis should be placed on increasing λ rather than μ.

Example – practical system augmentation for increasing ρ: We propose the bag structure of an augmented system, where an increase in resources (such as an improved resolution of the sensors, or an increased computational capability), can be allocated to include the increased set of traits, and trait-instances, as described in Table II, yielding an impressive ρ in the order of eighty million, which may be suitable for several applications.

TABLE II
AUGMENTED SET OF FACIAL SOFT BIOMETRIC TRAITS AND THE CORRESPONDING NUMBER OF TRAIT INSTANCES

Skin Color	**Hair Color**	**Eye Color**	**Glasses presence**	**Beard presence**	**Moustache presence**
3	8	6	2	2	2

Age	**Gender**	**Make up**	**Facial shapes**	**Facial feature shapes**
3	2	4	3	3

Facial measurements	**Facial feature measurements**	**Facial moles and Marks**	**Hair length**
3	6	6	3

This approach in turn, brings to the fore the issue that increasing ρ, may indeed result in an increased $E[\rho_e]$, but

might affect the correlation between the different categories. This would subsequently result in a reduced spread of Φ, which would imply a reduced distinctiveness in authentication. In regards to this, we give some intuition on the distinctiveness of some non-empty categories of the proposed system, by computing the correlation between these categories using Pearson's product-moment coefficient

$$r_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \quad (7)$$

The resulting correlation parameters shown below

$$r_{3,2} = -0.1964 \quad (\text{eye, hair})$$
$$r_{3,1} = 0.3770 \quad (\text{eye, skin})$$
$$r_{2,1} = -0.1375 \quad (\text{hair, skin})$$
$$r_{5,4} = 0.6359 \quad (\text{moustache, beard})$$

revealed as expected the highest correlation to be that between moustache and beard mirroring the fact that among the studied population the presence of moustache, given the presence of beard, is at 97.8%.

Further intuition on related correlations is given in Figure 5, which shows the joint pdf with respect to the eye, skin and hair color.

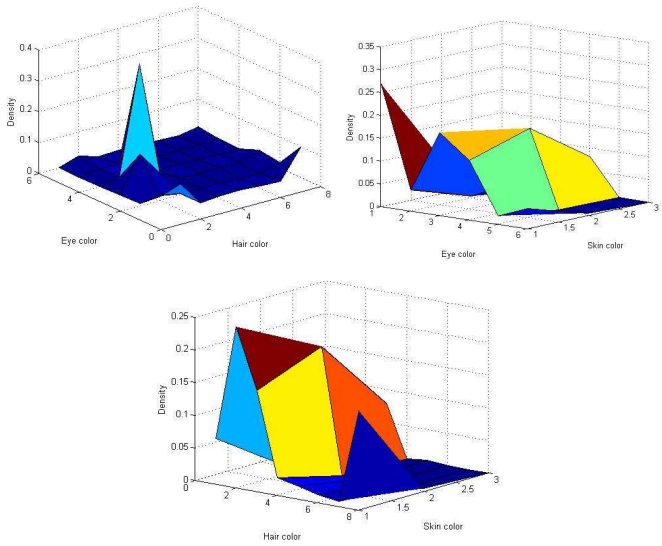

Figure 5. Color based soft biometric traits (eye – hair, eye-skin and hair-skin color) distributions for Feret database

B. Bounding N for a given interference probability

We are here interested in describing the relationship between N, and the corresponding probability of interference, as a function of Φ. We proceed to properly define the event of collision or interference.

Definition: The event of c*ollision*, or equivalently of *interference*, describes the event where *any* two or more subjects belong in the same category φ. Focusing on a specific subject, we say that this subject experiences interference if he/she belongs in a category which also includes other subjects from the authentication group.

In regards to this, we are interested in gaining insight on two probability measures. The first measure is the probability $p(N;\rho)$ that the authentication group of size N, chosen randomly from a large population of subjects, is such that there exist two subjects within the group that collide. We briefly note the relationship of $p(N;\rho)$ to the famous birthday paradox. For the other measure of system reliability, we consider the case where an authentication group of size N is chosen randomly from a large population of subjects, and where a randomly chosen subject from within this authentication group, collides with another member of the same group. We denote this probability as q(N), and note that clearly q(N) < p(N). To clarify, p(N) describes the probability that interference exists, even though it might not cause error, whereas q(N) describes the probability of an interference induced error.

We first focus on calculating and plotting p(N), under the simplifying assumption of statistical uniformity of the categories. The closed form expression for this probability is easily derived (see [14]) to be

$$p(N;\rho) = \begin{cases} 1 - \prod_{k=1}^{N-1}\left(1 - \frac{k}{\rho}\right) & N \le \rho \\ 1 & N > \rho \end{cases} \quad (8)$$

or equivalently expanded as

$$p(N;\rho) = 1 - \left(1 - \frac{1}{\rho}\right)\left(1 - \frac{2}{\rho}\right)\ldots\left(1 - \frac{N-1}{\rho}\right) = 1 - \frac{\rho!}{\rho^n(\rho-N)!} \quad (9)$$

We note that under the uniformity assumption, the above described $p(N;\rho)$ forms a lower bound on this same probability (in the absence of the same assumption). Equivalently, from the above, we can also compute the maximum N that will allow for a certain probability of collision. In terms of a closed form expression, this is accommodated by using the approximation from [15]:

$$p(N;\rho) \approx 1 - \left(\frac{\rho-1}{\rho}\right)^{\frac{N(N-1)}{2}} \quad (10)$$

and then solving for N to get

$$N(p;\rho) \approx \sqrt{2\rho \cdot ln\left(\frac{1}{1-p}\right)} \quad (11)$$

corresponding to the value of N for which the system will introduce interference probability equal to p. As an example, we note that for ρ = 1152, and p=0.5, then N = 39.

In regards to q(N), the closed form expression is readily seen to be

$$q(N) = 1 - \left(\frac{\rho-1}{\rho}\right)^{N}. \quad (12)$$

As an example we note that under the uniformity assumption, and given $\rho = 1152$, and q = 0.5, then N>700, which, as expected, is much higher than the pessimistic equivalent corresponding to $p(N,\rho)$.

Towards generalizing, we deviate from the uniformity assumption, to rather consider a more realistic setting where the category distribution originates from an online survey (see [11]), of 5142 subjects from the Central Germany region. For computational simplicity we choose to consider a simpler, reduced version of our proposed system, where the traits are limited to hair color and eye color. In this setting, the hair color trait has 7 trait-instances, and the eye color trait has 5 trait instances, resulting in a total of ρ=35 categories, with probabilities $P(\varphi_i)$, i=1,…, 35.

In this case the probability that all N subjects are in different categories is the sum of the products of all non-colliding events [13]:

$$p(N;\rho) = 1 - \sum_{\alpha\neq\beta\neq\cdots\neq\omega} \mathrm{P}(\varphi_\alpha)\mathrm{P}(\varphi_\beta)\ldots\mathrm{P}(\varphi_\omega) \qquad (13)$$

where the summation indexing corresponds to the non-empty categories with respect to the authentication group. This probability is plotted in Figure 6, where we note that as expected this probability exceeds the probability resulting under the uniformity assumption, albeit not by much.

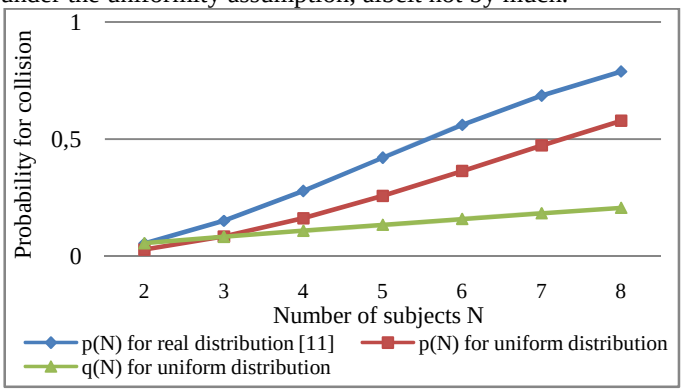

Figure 6. q(N) and p(N) for real and uniform distribution

C. *Simulation evaluation of the system in the interference limited setting of very high sensor resolution*

In the following we provide a simulation of the probability of identification error, in the setting of interest, under the assumption that the errors are due to interference, i.e., under the assumptions that errors only happen if and only if the chosen subject shares the same category with another person from the randomly chosen authentication group. This corresponds to the setting where the soft-biometric approach cannot provide conclusive authentication. In the simulation, the larger population consisted of 646 people from the Feret database, and the simulation was run for different sizes N of the authentication group. The probability of authentication error is described in the following figure.

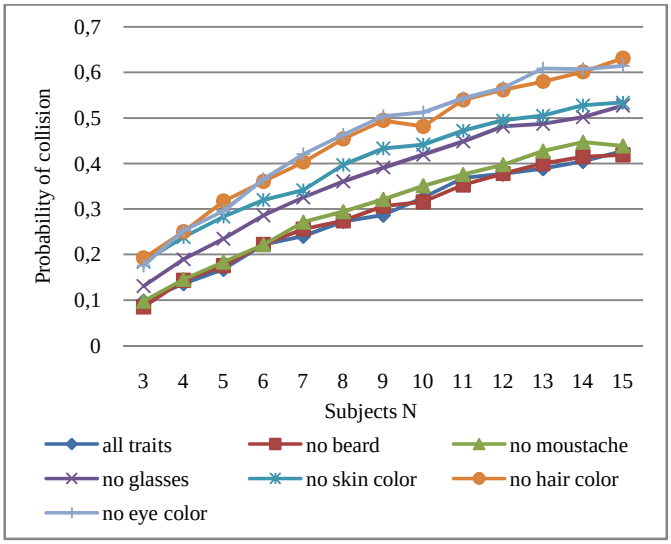

Figure 7. Collision probability in an N sized authentication group

As a measure of the importance of each trait, Figure 7 describes the collision probability when different traits are removed. The presence of moustache and beard seem to have the least influence on the detection results, whereas hair and eye color have the highest impact on distinctiveness.

V. Conclusions

This work introduces the novel idea of using a bag of facial soft biometrics for person verification and identification. The novel approach requires a novel framework, which was used to construct a specific soft-biometric system, for which parameters and algorithms were constructed to strike a proper balance between complexity and performance. Finally statistical analysis revealed related limitations, as well as future directions towards better understanding and improving these limitations.

Future work may include taking a closer look at the role of the specific detection algorithms in defining the overall system performance, as well as study of the of algorithmic dependencies, and identification of methods that would allow such dependencies to improve reliability and performance.

References

[1] Henry T.F. Rhodes, *Alphonse Bertillon: Father of Scientific Detection*, Abelard-Schuman, New York, Greenwood Press, 1956.

[2] A. K. Jain, S. C. Dass, and K. Nandakumar, "Soft Biometric Traits for Personal Recognition Systems," in *Proc ICBA*, 2004, pp. 731-738.

[3] A. K. Jain, S. C. Dass and K. Nandakumar, "Can soft biometric traits assist user recognition?" in *Proc. of SPIE*, 2004, vol 5404, pp. 561-572.

[4] X. Jiang, M. Binkert, B. Achermann, and H. Bunke, "Towards Detection of Glasses in Facial Images," *Pattern Analysis & Applications*, Springer London, vol. 3, pp. 9-18, 2000.

[5] (2010) Caltech Database [Online]. Available: vision.caltech.edu/html-files/archive.html

[6] (2010) The Feret website. [Online]. Available: face.nist.gov/colorferet/

[7] P. Viola and M. Jones, "Robust real-time face detection," in *Proc. ICCV*, 2001, pp. 747-747.

[8] P. Kakumanua, S. Makrogiannisa, and N. Bourbakis, "A survey of skin-color modeling and detection methods", *Pattern Recognition*, vol. 40, issue 3, March 2007.

[9] M. Zhao, D. Sun, and H. He, "Hair-color Modeling and Head Detection," in *Proc. WCICA*, 2008, pp.7773-7776.

[10] (2010) OpenCV webpage on Source forge. [Online]. Available: sourceforge.net/projects/opencvlibrary/

[11] (2010) Hair and eye colors webpage on haar-und-psychologie. [Online]. Available: haar-und-psychologie.de/haarfarben/hair-colors-eye-colors-germany-austria-switzerland.html

[12] S. Denman, C. Fookes, A. Bialkowski, S. Sridharan, "Soft-Biometrics: Unconstrained Authentication in a Surveillance Environment," in *Proc. DICTA*, 2009, pp. 196-203.

[13] K. Joag-Dev and F. Proschan, "Birthday problem with unlike probabilities" *American Mathematical Monthly*, vol. 99, Nr. 1, p.10-12, Jan. 1992.

[14] A. DasGupta, "The matching, birthday and the strong birthday problem: A contemporary review" *Journal of Statistical Planning and Inference*, vol. 130 (1-2), pp. 377-389, March 2005.

[15] S. E. Ahmed and R. J. Mcintosh, "An asymptotic approximation for the birthday problem," *Crux Mathematicorum*, vol. 26, pp. 151-155, Apr. 2000.

[16] (2010) Actibio website. [Online]. Available: actibio.eu