# Human Face Identification via Comparative Soft Biometrics

Nawaf Almudhahka          Mark Nixon          Jonathon Hare
School of Electronics and Computer Science
University of Southampton
Southampton, United Kingdom
{nya1g14, msn, jsh2}@ecs.soton.ac.uk

## Abstract

*Soft biometrics enable the identification of subjects based on semantic descriptions collected from eyewitnesses allowing people to search in surveillance databases. Although research has recently shown an increased interest in soft biometrics, not much of the work have used crowdsourcing, and it did not investigate the impact of feature selection on identification. In this paper, we introduce a new set of facial soft biometrics and labels with a novel description for the eyebrow region. Also, we examine the use of crowdsourcing for labelling the comparative facial soft biometrics and assess its impact on the identification. Moreover, we explore the impact of feature selection with our biometric measures and evaluate the effect of label scale compression. Experiments based on the Southampton biometric tunnel database demonstrate a 100% rank-1 identification rate using 20 features only.*

## 1. Introduction

Biometrics provide means of identifying people automatically based on their physical or behavioral traits. Traditionally, 'hard' biometrics, including fingerprints, DNA, and iris, have been used to identify people. However, hard biometrics require an individual's cooperation and can be computationally time consuming [1]. With the recent growth of surveillance applications that operate under adverse visual conditions such as long distance and low camera resolution [2, 3], a need arose for biometric attributes that enable the identification of humans under these conditions and which do not require the cooperation of the individuals. In addition, the difference in the way by which humans and machines identify people prevented the utilization of humans' description in automatic identification of people. All these factors have motivated the introduction of soft biometrics, which are the physical and behavioural traits that can be semantically described by humans [1-3]. The problem of human identification using soft biometrics has been studied by many researchers, and different forms of soft biometrics have been proposed such as: body, head, face, and clothing [1-10]. Among all aforementioned soft biometrics, facial traits are considered as the most informative and discriminative traits for human


Figure 1: Examples of wanted subjects in London riots 2011.[1]

identification at distance [8, 9]. Most of the research in human identification using facial soft biometrics has been based on using categorical labels [4, 6, 7], nonetheless, the use of comparative labels to describe people by their differences resulted in a more accurate representations of the traits and improved identification accuracy [2]. Moreover, previous studies on facial soft biometrics are limited in labeling the soft biometric on local groups, who annotate the traits of the datasets used in the studies [1-3,5], while the involvement of crowdsourced contributors in labelling the facial soft biometrics has not been well investigated. Also, the previous studies did not address the application of feature selection algorithms within facial soft biometrics and the implications of these algorithms on the identification performance.

This study sets out to assess the effect of crowdsourced comparative facial soft biometric labels on the identification performance, and to investigate the role of different feature selection algorithms in improving the identification performance when applied with comparative facial soft biometrics. The terms "trait" and "feature" are used synonymously throughout this study.

The major contributions of this study are:
- The definition of an enriched set of facial soft biometrics that is more appropriately described.
- A more complete use of eyebrow traits that emphasizes the eyebrow role in human face recognition [11, 12].
- The crowdsourcing of facial soft biometrics comparisons, which enables a more robust assessment of the traits and labels.

[1]BBC News: bbc.co.uk/news/uk-england-london-16171972

- The application of feature selection algorithms on the collected comparative labels.

Section 2 introduces the facial soft biometrics and comparative labels, Section 3 describes the data collection and analysis of the labels. Section 4 explains the experimental platform and reports the results with discussions. Finally, Section 5 summarizes the findings.

## 2. Facial Soft Biometric Traits

The human face has various traits that have different levels of distinctiveness and saliency. To consider a certain facial trait as a soft biometric identifier, it should retain an adequate level of discriminative power (i.e. be identifiable), and it should be easy to remember and describe by humans [8]. Although facial biometrics are vital in law enforcement, they lack for a vocabulary to describe them [3]. In this work, we have defined a new set of facial soft biometrics and the corresponding comparative labels. Our selection of traits has paid a special attention to the upper components of human face (i.e. eyebrows, and eyes) due to their significance in face recognition [11, 12]. One of the novelties of this work is the use of eyebrow vertical height and horizontal length as facial soft biometrics in addition to introducing a description of forehead hair. Other face components (e.g. nose, mouth, cheek, etc.) were included in our soft biometrics set assess their discriminative power.

All the labels used in this study are comparative since this has been shown to have greater discriminative capability than categorical description [2], except for *eyebrow shape* since it requires many details in its description. Also, the *global* soft biometrics (i.e. *gender*, *figure*, *skin colour*, and *age*) were included with the soft biometric set to assess their impact on the identification performance. The *gender* was in a comparative format as introduced in [13] to achieve a level of consistency among the traits. Table 1 shows the traits selected for this study along with the terms that are used as comparative labels. A trait comparison between two subjects is defined as the extent to which the trait differs between the two subjects and its value is set based on a five-point bipolar scale that ranges from -2 to +2 according to the label assigned to that trait. All the traits have a "Do not know" label to address the cases in which a trait is occluded. Nevertheless, the experimental results have shown that compression of the comparison levels improves the identification performance. Therefore, the five-point scale was compressed to a three-point scale by excluding the -1 and +1 levels from the rating and the inference (more insights are presented in Section 4).

## 3. Methodology

### 3.1. Data Acquisition Through Crowdsourcing

The dataset used for this study was extracted from the Southampton biometric tunnel database [14], which contains video recordings from more than 200 subjects using different viewpoints. The dataset was formed using the frontal video frame that was closest to the camera for 100 different subjects. The face of each subject was automatically detected on each frame using the Viola-Jones technique [15], then it is cropped from the rest of the scene. In a similar approach to that used in [16], all the face images were normalised by adjusting the inter-pupil distance to 70 pixels to ensure a rational comparisons between the subjects. Also, it is important to mention that the images used in this study have a low resolution and are affected by motion blur. In addition, some subjects exhibited a skewed face pose, and some have partial occlusion of some traits due to pose variability, long forehead hair, or presence of spectacles as shown in Figure 2. These factors make the dataset more reflective of the actual visual conditions of surveillance, which are characterised by: low resolution; high pose variability; and partial occlusion of faces.

The labelling of the facial soft biometrics of the subjects was accomplished through crowdsourcing. A total of 4950 subject comparison tasks were generated and launched as a job on the CrowdFlower platform as shown in Figure 3. For each crowdsourced task, the contributor was required to compare the facial traits between the two subjects using the labels listed in Table 1. The job resulted in the completion of 3522 comparisons whereas the remaining comparisons were inferred. The dataset of the collected labels will be publically available. Table 2 summarizes the job statistics.

### 3.2. Relative Ranking of Traits

During the labelling of facial soft biometrics, each comparison was given a label value that represents the difference in the trait between two subjects. There is an essential need to rate the subjects according to their traits' strengths in order to analyse the traits and labels defined in this study in addition to generate a feature vector for each subject, which is used in the identification. The Elo rating system [8] is a well-known scheme that is used to rank chess players based on their expected and actual scores. Given two players $A$ and $B$ with the rates $R_A$ and $R_B$ respectively, the expected score $E$ for each of the players is calculated as follows:

$$E_A = 1 \Big/ {1+10^{\frac{R_A - R_B}{400}}} \quad (1)$$

$$E_B = 1 \Big/ {1+10^{\frac{R_B - R_A}{400}}} \quad (2)$$

Based on the game outcome, the updated rates, $\bar{R}_A$ and $\bar{R}_B$, for subjects $A$ and $B$ respectively are:

$$\bar{R}_A = R_A + K(S_A - E_A) \quad (3)$$
$$\bar{R}_B = R_B + K(S_B - E_B) \quad (4)$$

where $K$ is the score adjustment factor and $S$ is the actual score of the player that is set as: 1 for a win; 0 for a loss; and 0.5 for a draw. The Elo rating system was used to generate the relative score of the subjects for each trait based on the strength of the trait presence in the subject. Each trait comparison between two subjects is considered as a game between two players, and a subject's relative rank for the trait is set based on the comparative label value, which ranges from -2 to 2. The relative rates of the trait are used to create the subject's feature vector that is used in the identification process (described in Section 4).

## 3.3. Data Analysis

The analysis was performed with the relative scores for each trait and has two objectives: exploring the normality

Table 1: Facial Traits and Corresponding Comparative Labels

| No. | Trait | Labels |
|---|---|---|
| 1 | Age | [Much Younger, More Young, Same, More Old, Much Older] |
| 2 | Figure | [Much Thinner, More Thin, Same, More Thick, Much Thicker] |
| 3 | Skin Colour | [Much Lighter, More Light, Same, More Dark, Much Darker] |
| 4 | Gender | [Much Feminine, More Feminine, Same, More Masculine, Much Masculine] |
| 5 | Eye Shape | [Much Tilted Inward, More Tilted Inward, Same, More Tilted Outward, Much Tilted Outward] |
| 6 | Eye Size | [Much Smaller, More Small, Same, More Large, Much Larger] |
| 7 | Inter-pupil Distance | [Much Closer, More Close, Same, More Wide, Much Wider] |
| 8 | Eye-to-Eyebrow Distance | [Much Smaller, More Small, Same, More Large, Much Larger] |
| 9 | Eyebrow Thickness | [Much Thinner, More Thin, Same, More Thick, Much Thicker] |
| 10 | Eyebrow Length | [Much Shorter, More Short, Same, More Long, Much Longer] |
| 11 | Inter Eyebrow Distance | [Much Closer, More Close, Same, More Wide, Much Wider] |
| 12 | Chin Height | [Much Smaller, More Small, Same, More Large, Much Larger] |
| 13 | Jaw Shape | [Much Chiseler, More Chiseled, Same, More Lantern-Shaped, Much Lantern-Shaped] |
| 14 | Cheek Shape | [Much Flatter, More Flat, Same, More Prominent, Much Prominent] |
| 15 | Cheek Size | [Much Smaller, More Small, Same, More Large, Much Larger] |
| 16 | Forehead Hair | [Much Less Hair, Less Hair, Same, More Hair, Much More Hair] |
| 17 | Mouth Width | [Much Narrower, More Narrow, Same, More Wide, Much Wider] |
| 18 | Lips Thickness | [Much Thinner, More Thin, Same, More Thick, Much Thicker] |
| 19 | Face Width | [Much Narrower, More Narrow, Same, More Wide, Much Wider] |
| 20 | Face Length | [Much Shorter, More Short, Same, More Long, Much Longer] |
| 21 | Nose Length | [Much Shorter, More Short, Same, More Long, Much Longer] |
| 22 | Nose Width | [Much Narrower, More Narrow, Same, More Wide, Much Wider] |
| 23 | Nose-Mouth Distance | [Much Shorter, More Short, Same, More Long, Much Longer] |
| 24 | Face Shape | [Much Ovoid, More Oval, Same, More Round, Much Rounder] |

Table 2. Crowdsourced job statistics

| | Collected | Inferred | Total |
|---|---|---|---|
| Total traits comparisons | 37968 | 78969 | 116937 |
| Total subjects' comparisons | 3522 | 1428 | 4950 |
| Average judgements per comparison | 2.22 | 33.68 | N/A |
| No. of subjects | 100 | | |
| Total number of contributers | 3073 | | |

for each trait, which aids in choosing the feature selection algorithm; and studying the correlation between the traits.

**3.3.1** *Normality Test:* To assess the normality of traits data, Anderson-Darling statistical test, was applied with the relative scores for each trait. Trait data is considered as normally distributed if the resulted *p*-value $p \geq 0.05$. The test showed that: *age*, *figure*, *skin colour*, *eyebrow to eye distance*, *eyebrow length*, and *forehead hair* violate the normality assumption. The normality violation for *age*, *figure*, and *skin colour* can be attributed to the database makeup, while the normality violation of *eyebrow length* and *eye to eyebrow distance* might be attributed to the high exposure of eyebrow to beauty treatments. Finally, the rejection of normality for *forehead hair* is likely caused by the high level of variability associated with front hair.

**3.3.2** *Trait Correlation:* To discover association between the traits, we have performed correlation analysis and the Pearson's correlation coefficient, *r*, was computed for all the possible pairs of traits. Figure 3 shows the correlations between the traits. The most significant correlations are found between: face measurement traits (e.g. *face length*, and *face width*) and structural traits of face (e.g. *face shape*, and *jaw shape*); *figure* and *face width*; in addition to *eyebrow thickness* and *gender*. In general, these results show the independence of most traits, which contributes to the discriminating power of these trait.

## 3.4. Feature Selection

One of the main contributions of this work is the derivation of an optimised feature set that reduces the number of features used in identification while improving (or maintaining) the identification performance. Wherefore, we have used three different feature selection algorithms to measure the discriminating power of the traits: mutual information [17]; Analysis of Variance (ANOVA) [18]; and Kruskal–Wallis [19]. In this section, we present the features ranking according to each of the three algorithms, while their impact on identification is demonstrated in Section 4.

**3.4.1** *Mutual Information:* Mutual information reveals the amount of information carried by one random variable, *X*, about another random variable, *Y*. Mutual information was computed for each of the traits illustrated in Table 1 according to the following formula:

$$MI(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right) \quad (5)$$

where *X* represents the trait relative scores and *Y* represents the subjects' labels, *p(x)* and *p(y)* are probability mass functions for the variables *X* and *Y* respectively, and *p(x,y)* is the joint probability mass function.

**3.4.2** *Analysis of Variance (ANOVA):* The one-way ANOVA test is based on $F$ statistic. The greater is the $F$ ratio, the stronger is the discriminative capability of the trait. The $F$ ratio is calculated for each of the traits in this study as follows:

$$F = \frac{\sum_i n_i (\overline{X_i} - \overline{X_G})^2 / (k-1)}{\sum_{ij} (X_{ij} - \overline{X_i})^2 / (N-k)} \quad (6)$$

where $N$ is the total number of subjects, $k$ is the number of groups, $\overline{X_i}$ is the mean relative score of the $i^{th}$ group, $\overline{X_G}$ is the overall mean of the relative scores of the $k$ groups, $X_{ij}$ is the relative score for the $j^{th}$ subject in the $i^{th}$ group, and $n_i$ is the number of subjects in the $i^{th}$ group.

**3.4.3** *Kruskal-Wallis Analysis of Variance:* Kruskal-Wallis test is the non-parametric equivalent of one-way ANOVA. It was used to address the non-normality of some traits. The Kruskal-Wallis test is based on the $H$ statistic. Larger values of the $H$ statistic imply stronger discriminative capability. The $H$ ratio is calculated for each trait as follows:

$$H = (12/(N(N+1))) \sum_{i=1}^{a} n_i \left( \overline{R_i} - (N+1)/2 \right)^2 \quad (7)$$

where $a$ is the number of populations, $n_i$ is the number of subjects per population, $N$ is the sum of population sizes, and $\overline{R_i}$ is the average rank of the $i^{th}$ population.

The results of applying each of the feature selection algorithms are illustrated in Figure 5. The results show that *skin colour* has the highest discriminating power, followed by *eyebrow length*, which is a novel contribution of this work. Also, the traits derived from mouth region (i.e. *lips thickness*, *mouth width*, and *nose-to-mouth distance*) in addition to some of the traits derived from eye and eyebrow region (i.e. *eye-to-eyebrow distance*, *inter-eyebrow distance,* and *eyebrow thickness*) have shown a notable discriminating ability, which support our emphasize on these traits. With respect to the *global traits* other than *skin colour* (i.e., *age, figure,* and *gender*), they have also shown
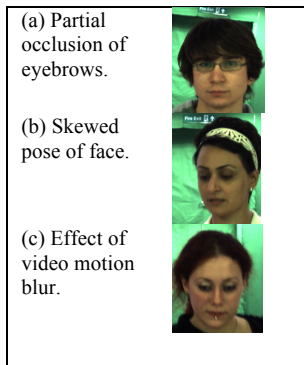

Figure 4. The correlations between the traits
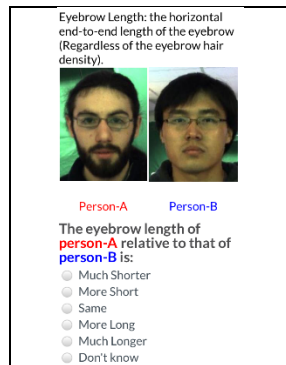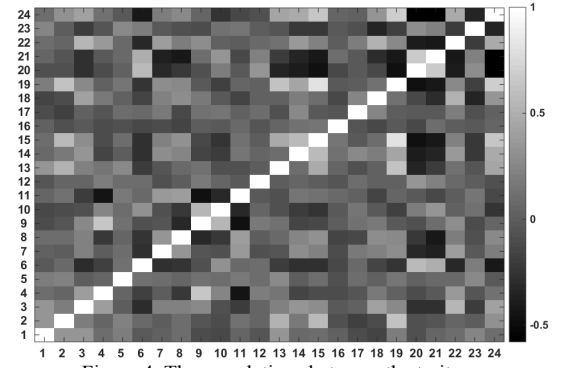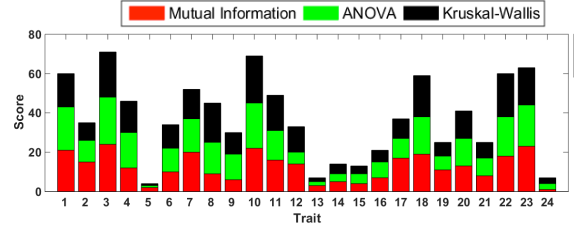

Figure 5. Soft biometric traits scores based on the aggregated ranking from the three feature selection algorithms.

a substantial discriminating power. The results also indicate that *cheek size* and the shape-based traits, such as: *cheek shape,* and *face shape*, have the lowest discriminating power compared to the other traits. Contrary to our assumptions, the results show that the *inter-pupil distance* has a significant discriminating power, although it was set to 70 pixels for the whole dataset. A possible explanation for this is that the annotators' perception for *inter-pupil distance* could be affected by other traits such as: *eye size*, *eye shape*, and *face measurements*.

## 4. Identification Using Comparative Facial Soft Biometrics

The purpose of human identification using facial soft biometrics is to recognize an unknown subject based on a semantic description of the subject's facial features. There are three objectives behind performing human identification using comparative facial soft biometrics in the context of this research: (1) assessing the impact of our new set of facial soft biometrics and their comparative labels on the identification performance; (2) exploring the implications of applying the three feature selection algorithms (i.e. mutual information, ANOVA, and Kruskal Wallis) on identification; and (3) determining the effect of labels scale compression on identification. To meet the aforementioned objectives, an identification experiment was designed and performed over all the 100 subjects that constitute the dataset used in this study with difference number of comparisons and using different features subsets. The performance estimation was based on Leave One Out Cross Validation (LOOCV). For each subject *s* of
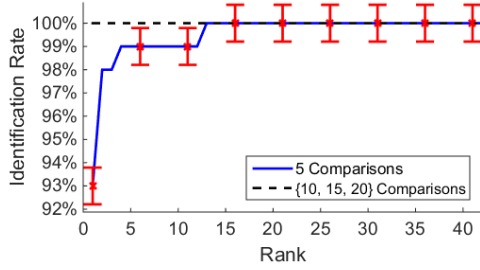


(a) Partial occlusion of eyebrows.

(b) Skewed pose of face.

(c) Effect of video motion blur.

Figure 2. Samples of some face images.



Eyebrow Length: the horizontal end-to-end length of the eyebrow (Regardless of the eyebrow hair density).

Person-A          Person-B

**The eyebrow length of person-A** relative to that of **person-B** is:
- Much Shorter
- More Short
- Same
- More Long
- Much Longer
- Don't know

Figure 3. The crowdsourced job launched to collect labels.

Figure 6. Identification performance using the full feature set.


Figure 7. The effect of the feature selection on the identification performance using: (a) 5, (b) 10, (c) 15, and (d) 20 comparisons.

the $N$ total subjects of the dataset, $c$ counterpart subjects are randomly selected and the corresponding comparison between subject $s$ (the probe) and each of the $c$ counterpart subjects is excluded from the global comparisons set $M$. Then, the relative score, $t$, for each element, $k$, of the trait set $T$ (illustrated in Table 1) for the $N$ subjects is generated using Elo rating system based on the $M$-$c$ remaining comparisons. Similarly, the relative scores for the traits of the probe $s$ are generated based on the $c$ comparisons excluded earlier from the set $M$. The sum of the Euclidean distances $D$ between the probe $s$ and each subject in the dataset $i$ is computed as follows:

$$D(s,i) = \sqrt{\sum_{k=1}^{T} \left( t_k(s) - t_k(i) \right)^2} \qquad (8)$$

and the $N$ subjects are ranked based on their corresponding $D$ in an ascending order where the subject that have the minimum $D$ is considered as a rank-1 match with the probe $s$. This experiment was performed for each subject with 5, 10, 15, and 20 comparisons using different feature subsets that were generated using mutual information, ANOVA, and Kruskal Wallis. The overall experimental procedure was repeated 30 times and the mode of the 30 trails was considered as the outcome. Cumulative Match Characteristic (CMC) is generated based on the position of the probe $s$ in the rank and used as a performance metric.

## 4.1. Identification Using Facial Comparative Labels

The first experiment aimed to measure the identification performance of the new soft biometrics and labels set defined in this study following the experimental procedure described in the introduction of this section. The experiment included all the traits listed in Table 1 (i.e. the full feature set) and it was performed with different number of comparisons. The results are shown in Figure 6. The most striking result to emerge from this experiment is the achievement of rank-1 identification rate of 100% using 10 comparisons only. This result outperforms the result of a prior study by Reid [3] in three respects: (1) the number of comparisons required to achieve the 100% rank-1 identification rate is 10, while the same rate was achieved with 20 comparisons in [3]; (2) a slightly smaller number of features (24) was used in this study as compared to [3],
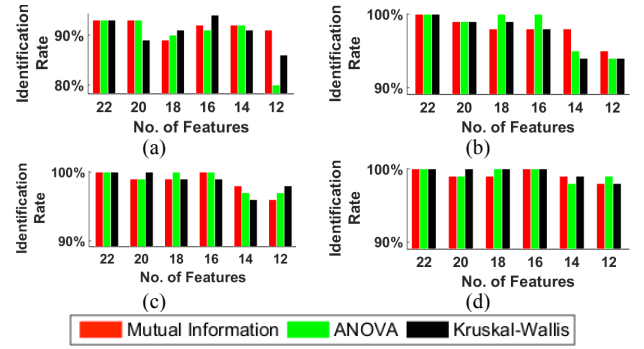
which was based on 27 features, and this indicates the effectiveness of our new soft biometric set; (3) the labelling in this work is based on frontal face images only, while the experiment in [3] depended on frontal and side face images, and this shows the efficiency of crowdsourcing in labelling the facial features. Moreover, these results outperform the identification results achieved by augmenting body with clothing comparative labels in [5] in which a 100% identification rate was achieved at rank 30.

## 4.2. The Effect of Feature Selection

In section 3, we have presented three feature rankings based on three feature selection algorithms: mutual information [17]; ANOVA [18]; and Kruskal Wallis [19]. To analyse the impact of these algorithms on the identification performance, a series of identification experiments was performed using feature subsets that were derived from each of the three rankings with different number of comparisons. Figure 7 shows the identification performance at rank-1 for each subset and we can see that the impact of each subset is highly dependent on the number of comparisons used in the identification in addition to the subset size. The most significant differences in impact are observed when 5 comparisons are used. As the comparisons increases to 10, the subsets derived from ANOVA ranking have the best performance impact among the other subsets for the larger sizes (i.e. 18 features and above), thus the 100% identification rate was achieved by the top 18 and 20 features derived from ANOVA ranking. Moreover, these two subsets maintain their superiority as the number of comparisons increases to 15 and 20 thereafter. As the number of features used in the identification decreases below 18, the subsets derived from mutual information ranking show the best performance impact although the maximum identification rate is not maintained at 100%. These results demonstrate the effectiveness of our facial soft biometrics set as 18 features are sufficient to achieve the maximum identification rate of the full feature set. The results also reveal the effectiveness of ANOVA for feature selection in the context of this study. Furthermore, the
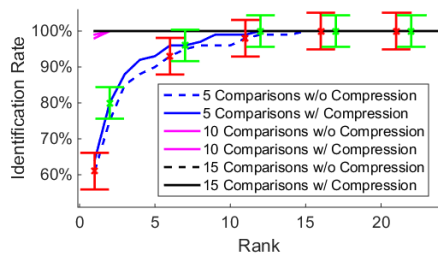
Figure 8. The effect of label scale compression on identification.

results suggest that Kruskal-Wallis algorithm has no significant advantage over ANOVA for our soft biometrics, and this may be due to the fractional effect of non-normality, which is presented in 6 out of the 24 features. Nevertheless, it does not reduce the importance of assessing the features normality prior to choosing a feature selection algorithm.

### 4.3. The Effect of Label Scale Compression

As mentioned in section 3, the experimental results showed a significant increase in the identification rate when the five-point bipolar scale is compressed to a three-point scale by excluding the levels of -1 and +1 (i.e. the "less" and the "more" labels). Figure 8 shows the performance gains resulted from an identification experiment that was performed to assess the effect of scale compression. A possible explanation for this gain is the insignificance of levels -1 and +1 in distinguishing the traits of the subjects, which is caused by the subjectivity of visual perception by humans [20]. Another possible explanation is that the scale compression to three levels is more reflective of the actual scenario assumed by the Elo rating system, which is based on three game outcomes (i.e. win, loss, or draw). Also, this gain could be attributed to the effect of introducing categorical description to the comparative basis, which is antithetical to the nature of comparative descriptions. Certainly, reducing the comparisons to greater or less than and the same, improves performance considerably.

## 5. Conclusions and Discussion

The purpose of this study was to introduce a enriched set of facial soft biometrics and to determine the effect of crowdsourced comparative labels on the identification performance, in addition to evaluating the impact of feature selection on the identification performance. The experiments showed that the use of our new facial soft biometrics set and the crowdsourcing of soft labels improve the identification performance that a rank-1 identification rate of 100% was achieved using 10 comparisons only. Also, the experiments revealed that feature selection could reduce the number biometric measures used in identification while preserving the identification rate. In addition, we have shown the positive effect of soft labels scale compression on the identification performance. Taken together, the findings of

this study highlight the effectiveness of our new facial soft biometrics set as well as the role of crowdsourcing in improving the human identification performance using comparative labels, and it show the potential gains from applying feature selection on facial soft biometrics. The dataset of the crowdsourced annotations will be publicly available to enable further studies on the collected labels.

## References

[1] Reid, D. A., Samangooei, S., Chen, C., Nixon, M. S., & Ross, A.. Soft biometrics for surveillance: an overview. *Machine learning: theory and applications. Elsevier*, 2013.

[2] Reid, D. A., & Nixon, M. S. Using comparative human descriptions for soft biometrics. In. *IEEE IJCB*, 2011.

[3] Reid, D., & Nixon, M. S. Human identification using facial comparative descriptions. In *IEEE ICB*, 2013.

[4] Klare, B. F., Klum, S., Klontz, J. C., Taborsky, E., Akgul, T., & Jain, A.K. Suspect identification based on descriptive facial attributes. In *IEEE IJCB,* 2014.

[5] Jaha, E. S., & Nixon, M. S. Soft biometrics for subject identification using clothing attributes. In *IEEE IJCB,* 2014.

[6] Kumar, N., Berg, A. C., Belhumeur, P. N., & Nayar, S. K.. Attribute and simile classifiers for face verification. In *ICCV, 2009*.

[7] Tome, P., Fierrez, J., Vera-Rodriguez, R., & Nixon, M. S. (2014). Soft biometrics and their application in person recognition at a distance. *Information Forensics and Security, IEEE Transactions on*, *9*(3).

[8] Reid, Daniel. *Human identification using soft biometrics*. PhD Dissertation. University of Southampton, 2013.

[9] Han, H., & Jain, A. (2014). Age, gender and race estimation from unconstrained face images. *Dept. Comput. Sci. Eng., Michigan State Univ., East Lansing, MI, USA, MSU Tech. Rep.(MSU-CSE-14-5)*.

[10] Jain, A. K., & Park, U. Facial marks: Soft biometric for face recognition. In *ICIP IEEE,* 2009.

[11] Sadrô, J., Jarudi, I., & Sinhaô, P. (2003). The role of eyebrows in face recognition. *Perception*, *32*(3), 285-293.

[12] Davies, G., Ellis, H., & Shepherd, J. (1977). Cue saliency in faces as assessed by the" Photofit" technique. *Perception*, (6), 263-9.

[13] Martinho-Corbishley, D., Nixon, M. S. & Carter, J. N. Soft Biometric Recognition from Comparative Crowdsourced Annotations. In *ICDP*, 2015.

[14] Shutler, J. D., Grant, M. G., Nixon, M. S., & Carter, J. N. (2004). On a large sequence-based human gait database. In *Applications and Science in Soft Computing* (pp. 339-346). Springer.

[15] Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, *57*(2), 137-154.

[16] Klare, B., & Jain, A. K. (2010, September). On a taxonomy of facial features. In *IEEE Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth International Conference on* (pp. 1-8). IEEE.

[17] Guo, B., & Nixon, M. S. (2009). Gait feature subset selection by mutual information. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, *39*(1), 36-46.

[18] Veres, G. V., Gordon, L., Carter, J. N., & Nixon, M. S. What image information is important in silhouette-based gait recognition?. In *Proc. CVPR, 2004*.

[19] Ali Khan, S., Hussain, A., Basit, A., & Akram, S. Kruskal-Wallis-Based Computationally Efficient Feature Selection for Face Recognition. *The Scientific World Journal*, *2014*.

[20] Sánchez, D., Chamorro-Martınez, J., & Vila, M. A. (2003). Modelling subjectivity in visual perception of orientation for image retrieval. *Information processing & management*, *39*(2), 251-266.