On the Potential for Facial Attractiveness as a Soft Biometric

Abstract. This paper describes the first study on whether human facial attractiveness can be used as a soft biometric feature. By using comparative soft biometrics, with ranking and classification, we show that attractiveness does have the capability to be used within a recognition framework using crowdsourcing, by using groups from the LFW dataset. In this initial study, the Elo rating system is employed to rank subjects' facial attractiveness based on the comparative descriptions. We will show how facial attractiveness attributes can be exploited for identification purposes and can be described in the same way and can add to performance of comparative soft biometrics attributes. Attractiveness does not appear to be as powerful as gender for recognition. It does however increase recognition capability and it is interesting that a perceptual characteristic can improve performance in this way.

Keywords: Comparative Soft Biometrics, Face Recognition, Facial Attractiveness, Facial Attributes, Ranking.

1 Introduction

Society requires human identification for business as usual, and for security. Biometrics is the science of recognizing individuals based on physical features and attributes; these attributes are chosen from human characteristics that use unique data to perform identification. Soft biometrics are attracting a lot of interest with the spread of surveillance systems, and have been developed to augment the performance achieved with 'hard biometrics' [1, 2, 3].

Soft biometrics for identification relate to behavioral and physical attributes that can be semantically defined and are describable by other people [4]. Such traits can be used for recognition, which can be extracted in the form of descriptors, labels and measurements [5]. It relies on individual characteristics, which must often be used to identity persons at a distance and under adverse visual surveillance conditions [3, 6]. There are two kinds of soft biometrics for identification: comparative and categorical. 'Comparative' means a person's attributes are classified in comparison with those of another person [7]. For an annotator to judge a figure comparatively to another figure is more natural than using absolute relations 'categorical' [3].

Developments in closed-circuit television (CCTV) networks, which have expanded in recent decades world-wide, have enlarged the reliance on – and capacity of – surveillance data for identification. There remains considerable uncertainty, however, in the available data. This is a major challenge encountered by law enforcement organizations in criminal investigations. This challenge has motivated research in soft biometrics for identification, including the face [2, 3, 5, 8], body [9, 10], and clothing [11]. These studies have investigated the sources of attributes that permit human identification. Of these, the human face appears the most potent for recognition at short distances [6].

A feature of soft biometrics is that information is required to be clear, unforgettable, and describable for identification, to allow search of a dataset during eyewitness statements and identification. A key characteristic distinguishing unforgettable faces from less memorable ones is beauty or 'attractiveness', as many experimenters have noted [12]. This suggests that it is prudent to study facial attractiveness as a soft biometric. In Fig. 1 (as found in our study), one can differentiate the two faces by attractiveness, in the same way as by gender. The question then is whether attractiveness is a generic description that can be used to aid identification within a larger pool of subjects.

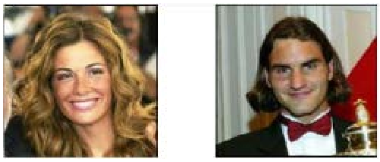

Fig. 1. Illustrating attractiveness as a face description.

Although numerous studies have investigated the impact of facial features on performance in recognition, no previous work has considered facial attractiveness as a soft biometric. Our contribution in this work is to investigate whether facial attractiveness can be considered a comparative soft biometric attribute and thus to contribute to face recognition. This paper appears to be the first to propose the use of attractiveness, to investigate the effects of facial attractiveness on facial recognition, and to investigate the accuracy of face recognition when attractiveness is included.

1.1 Importance of Facial Attractiveness

The attractiveness of a face plays a significant role in many daily and public activities. Evaluating or codifying facial attractiveness is, however, a very challenging concept, and has been debated by philosophers, artists, and scientists for many years [13]. The attractiveness of a face can apparently be easily decoded by the human brain, and can be hard to develop automated methods to measure it. Though attractiveness can be recognized instantly, it is still challenging to articulate how this can be achieved or even described [14].

Several studies have tried to use computational methods in order to analyze and evaluate facial attractiveness [13]. Although the study of facial attractiveness by a computational framework is still relatively new, it has potential for significant impact. Attractiveness has been the subject of significant consideration in the field, with the most common concentration being on integrating techniques from image processing, computer vision and machine learning. Multiple events highlight the capabilities and effectiveness of information inferred from facial attractiveness for recognizing people [14, 15].

In this work, a new set of proposed semantic facial attractiveness attributes is introduced, along with their comparative labels. We have discussed the background related to the problem. The remainder of the paper is organized as follows. Section 2 defines the new semantic facial attractiveness attributes and descriptive labels for annotation. Section 3 involves the crowdsourcing of comparative labels using the LFW dataset. Subsequently, it presents and defines the Elo rating system to compare features, for inferring relative measurements in the future. In section 4, the attractiveness attribute is analyzed via the correlation attributes and via feature selections. In section 5, the performance of the attractiveness attribute is evaluated, through the responses received for a sample of different subjects. Finally, the conclusions and suggestions for future work are put forward in section 6.

2 Attribute and Label Derivation

Recent psychological studies have shown that the human perception of facial attractiveness is largely constituted by variations between different individuals, for both male and female observers [16]. Moreover, findings have shown that estimations of beauty are dependent on the physiognomy of the face, and persons in all places use related criteria in their decisions. A strong relationship between beauty and specific features is presented in [17], which characterized neonate, mature and expressive attributes, such as a small nose, high forehead, prominent cheekbones or arched eyebrows. They concluded that the attractiveness ought not to be considered a poetic, inexpressible feature lying only in the eye of the beholder [14, 18].

Research shows that general attributes such as facial symmetry, averageness, and secondary sex characteristics are attractive in both men's and women's faces, and across cultures [14, 19, 20]. Nevertheless, the published studies have not yet reached a consensus on these hypotheses or rules [15]. To consider facial attractiveness features as soft biometric identifier, people should find them easy to remember and describe. Based on previous works on soft biometrics, a human identification technique is presented in [8] using comparative facial soft biometrics – creating a set of soft biometric attributes, which cover the most important facial components. In our study, we propose a list of 15 attributes similar to those used in a previous study. Seven more attributes are also added here as these might putatively relate to and describe facial attractiveness.

2.1 Label Comparisons of Facial Features

Naturally, people describe humans using estimations of physical attributes and labels. Comparative labels characterize the grade of comparisons of relative features. This study will use comparative labels associated with the attributes, defined based on a 4-point bipolar scale following a consistent format: "More A", "Same","Less A/ or More B" corresponding to label values 1, 0, and -1, respectively with -2 for "Can not See"

[8]. We need to measure the strengths fairly, because by establishing numerical value scores we can form an ordered list for each trait, and use this result to rank subjects. We have used attributes (listed in Table 1) of the most dependably understood facial attractive 'scales'. For each feature, an appropriate group of comparative labels is defined to describe these attributes.

Table 1. Soft facial attractiveness biometrics attributes and possible associated response labels.

	Soft Traits	**Comparative Labels**			
		1	**0**	**-1**	**-2**
1	Age	More old	Same	More young	Can not see
2	Attractiveness	Less Attractive	Same	More attractive	Can not see
3	Cheek shape	More flat	Same	More prominent	Can not see
4	Chin length	More long	Same	More short	Can not see
5	Eyebrow length	More long	Same	More short	Can not see
6	Eyebrow thickness	More thick	Same	More thin	Can not see
7	Eyes size	More large	Same	More small	Can not see
8	Face length	More long	Same	More short	Can not see
9	Face width	More wide	Same	More narrow	Can not see
10	Facial hair	Less facial hair	Same	More facial hair	Can not see
11	Forehead hair	Less forehead hair	Same	More forehead hair	Can not see
12	Gender	More masculine	Same	More feminine	Can not see
13	Jaw size	More narrow	Same	More wide	Can not see
14	Lip thickness	More thick	Same	More thin	Can not see
15	Nose length	More long	Same	More short	Can not see
16	Nose width	More wide	Same	More narrow	Can not see
17	Nose-mouth distance	More Short	Same	More Long	Can not see
18	Proportions	More average	Same	Less average	Can not see
19	Figure (Shape)	More fat	Same	More thin	Can not see
20	Skin color	More dark	Same	More light	Can not see
21	Skin smoothness	Less smooth	Same	More smooth	Can not see
22	Symmetry	Less symmetrical	Same	More symmetrical	Can not see

3 Deriving Facial Attractiveness Labels

3.1 Facial Attractiveness Dataset

We use the Labelled Faces in the Wild (LFW) dataset [21], which displays a large variety of the subjects. The LFW dataset is a popular database for studying unconstrained recognition and has been chosen since it reflects the environment of surveillance situations. Such variations more accurately reflect the challenges of image recognition in the real world. The aim here is to assess the general usage of attractiveness and collect physically unconstrained data to more accurately model normal human observations and explanation given to others, compared to additional limiting annotation jobs, as seen in [4].

In this study, we investigate the facial attractiveness of humans via comparative soft biometrics using subjects. A subset of subjects has been chosen within the dataset, in the total of 100 subjects, and four samples for each subject as the initial test for results in our study. In other words, these samples were used as justification for testing and computing the accuracy.

3.2 Crowdsourcing of Facial Attractiveness

Crowdsourcing is a relatively novel and effective method of collecting labels from human annotators for a dataset, and it is increasingly changing how datasets are created and deployed. A well-designed crowdsourcing job can be directed at a huge collection of high-quality comparative annotations and permits the collection of data from annotators from diverse national, linguistic and cultural backgrounds. This creates original ground-truth information and produces the chance to eliminate social annotation bias. Nowadays, with platform facilities, it is possible to assign annotation jobs to hundreds or more of computer-literate employees and produce verifiable outcomes in a matter of hours [22].

To construct and run the crowdsourced annotation job for this study, we used the Figure-Eight platform to gather labels. A worldwide network of contributors is connected to that platform; therefore, the questions must be clear and definitive. In addition, the platform offers control tools and widespread data analysis, and allows customers to agree to take a range of answers whilst refusing non-honest responses. Each labeler compared chosen attributes between two faces, using the labels in Table 1. An example of our crowdsourced comparison via the Figure-Eight platform is presented in Fig. 2.

Fig. 2. Example question from our Figure-Eight job.

For the chosen comparative labels, the annotation procedure required a user to compare one subject on the left, with another subject on the right. The style of each question is fundamentally a psychometric technique. Within the platform, test questions are designed and presented to the labelers to allow us to quantify the accuracy of the contributor and reduce the number of false answers. The crowdsourcing task was repeated to make sure that, by involving as many labelers as possible within our limited time frame, the responses could be regarded as strongly and confidently as possible. A total of 8,378 labelers, provided a total of 184,316 labels for the 400 subject images.

3.3 Ranking by Relative Attractiveness Attributes

All comparisons provided a label value as mentioned in section 2. The value of the label is used to calculate the comparative strength of the subjects' attributes, based on the various ranking systems. This will represent the difference in the feature between two subjects where complete or absolute measurements of the attributes cannot be detected due to the inexactness of verbal or written descriptions. Subsequently, to evaluate the labels and performance of facial attractiveness, it is necessary to rate or "rank" the subjects according to the powers or levels of their attributes. Various ranking algorithms such as the Elo ranking system can be applied for this purpose [3].

Elo is a well-known score-based rating system for rating the players of chess matches based on their estimated and real scores. Such a rating system could be used to rank subjects in a soft biometrics framework, including facial attractiveness as a comparative attribute. Its effectiveness for comparative soft biometrics has previously been confirmed and demonstrated in [3], where it was used to create biometric signatures and to determine the comparative rates of the features from comparative labels.

4 Facial Attractiveness for Recognition

To express a thorough understanding of the collaboration and association between facial attractiveness and other facial soft biometrics attributes, an advance analysis may be needed to investigate how these attributes combine for purposes of identification.

4.1 Correlation Analysis

To discover any dependences or relations between the attractiveness and the other 20 facial attributes, correlation was calculated for all the pairs of attributes. The analysis was performed with the relative scores for the features [15]. The correlation is performed using the Pearson's correlation coefficient, r which can be calculated as:

$$r = \frac{\sigma_{XY}}{\sigma_X \, \sigma_Y} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad (1)$$

Where X and Y are two variables representing the label values of two different semantic traits used to describe an individual. Specifically, X_i and Y_i are two labels of the i^{th} annotation describing the same subject. Thus, r is calculated by dividing the covariance of X and Y, represented as σ_{XY}, by the product of σ_X and σ_Y the standard deviations of X and Y correspondingly. The correlation matrix is calculated by using equation (1) is illustrated in Fig. 3.

To represent the correlation coefficient between two labels, the resulting value of r lies between 1 and -1. A value r = 0 means there is no linear correlation between labels. The nearer r is to 1 the more it reproduces a total higher positive linear correlation between labels. The nearer r is to -1 the more it reflects a total higher negative linear correlation between labels. From this figure, it is evident that there is a strong correlation between attractiveness and figure, proportions, smoother face skin, and symmetry (shapeliness). This supports what was previously discussed in section 2, about traits associated with facial attractiveness in most studies. In addition, we can determine some important positive linear correlations between attractiveness on one side, and two of global soft biometrics traits (age and gender) on the other [4, 9].

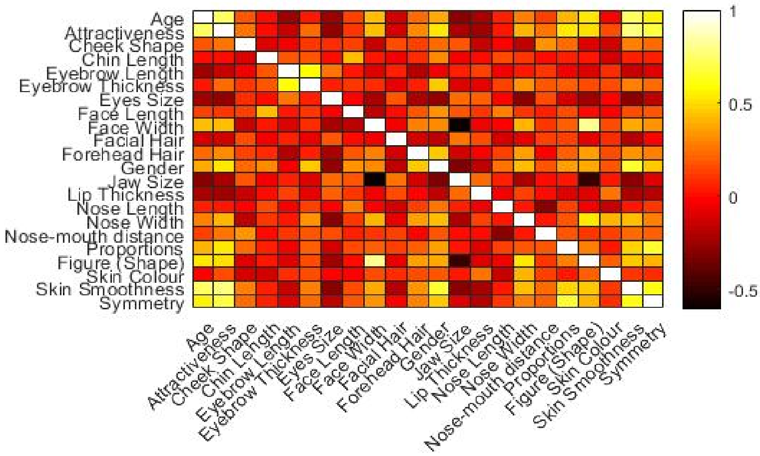

Fig. 3. Correlation between attractiveness and the other 21 facial attributes.

It can be seen from Fig. 4 that the results agree with our original intuition. For example, there is a positive linear correlation between attractiveness and between symmetry, smoother face skin and age; which means youngest people with smoother skin and more symmetry faces are very likely to be rated attractive, or vice versa, as shown in Fig. 4. The correlations and exploration of features' significance resulted in a better comprehension of the contribution of each attribute to identification and recognition, and may lead to improved or expanded predictability of other attributes. It is noted that supplementary investigations of correlations between couples of attributes, including attractiveness, could supply valuable information within a future study.

Fig. 4. The top images represent the subjects ranked top by attractiveness, while the bottom images show the lowest-ranked

4.2 Discriminative Power of Attractiveness

The estimation of an attribute's importance is significant for recognizing its strength as a semantic descriptor and to discover its contribution to human identification and recognition.

Mutual Information (MI): was used to evaluate the significance of the attractiveness attribute [23]. MI (typically written as $I(X, Y)$) is a quantity that measures how much a random variable, X, is related to another random variable, Y , (or vice versa). The MI is calculated as follows:

$$MI = I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \, ln \left(\frac{p(x, y)}{p(x) \, p(y)} \right) \qquad (2)$$

In the context of the facial attributes in a soft biometric framework, X represents the attribute relative rates and Y represents the subjects' labels, resulting in a computation of the MI of these two traits. While $p(x, y)$ is the joint probability distribution function for the random variables X and Y respectively, where $p(x)$ and $p(y)$ are the marginal probability distribution functions respectively.

In this study, MI was used to determine the discriminative power of the attractiveness attributes. MI was applied to the comparative rating data of each attribute, as shown in Fig. 5. It was computed for each of the 22 attributes discussed in section (2), according to equation (2).

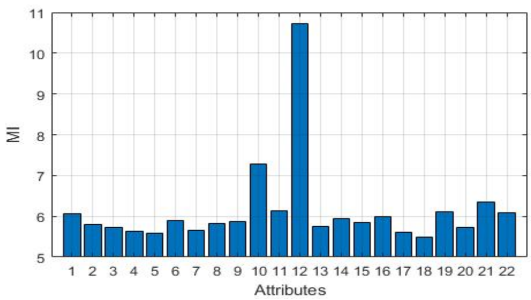

Fig. 5. Mutual information (MI) for each 22 attribute.

We can see that MI method rated attractiveness together with skin color as having modest discriminative power. Interestingly, some attributes such as gender, facial hair, skin smoothness, figure and symmetry were rated with high discriminative power. The results also show that proportions, eyebrow length and nose-mouth distance have the lowest discriminatory power compared with the other attributes. MI scoring methods enforce ranking for each trait independently. We can consider the attractiveness attribute to a discriminative attribute of the human face.

Sequential Floating Forward Selection (SFFS) Algorithm: is a familiar and widely-used trait subset selection method. It does not implement scoring for each feature individually, in contrast with the MI scoring method. Basically, SFFS is a bottom-up search technique which attains new attributes by starting with an empty set, then, at each repetition retaining the next best informative feature to the subset of selection. At each iteration, we form a soft biometric verification system, and change the specific set of features used, selecting the following attribute depending on the already selected features. After each forward step insertion of the best feature, SFFS implements backward rejection steps of the worst feature, as long as the new subset (after the elimination) increases the previous performance of a subset. This is done until the highest likely performance is achieved [24].

This analysis leads to practical understanding of how appropriate and significant attractiveness is to human face identification. By employing the SFFS algorithm, the attractiveness is listed within the top attributes by SFFS ordering of the soft biometrics attributes. Table 2 provides different trait ordering lists obtained by feature selection method.

Table 2. Ordered list of comparative facial soft traits inferred using SFFS methods.

Ordering	SFFS	Ordering	SFFS
1	Gender	12	Figure (Shape)
2	Facial hair	13	Nose-mouth distance
3	Age	14	Chin length
4	Attractiveness	15	Skin color
5	Forehead hair	16	Face width
6	Eyebrow thickness	17	Symmetry
7	Nose length	18	Eyebrow length
8	Skin smoothness	19	Eyes size
9	Lip thickness	20	Proportions
10	Jaw size	21	Face length
11	Nose width	22	Cheek shape

The attractiveness attribute is shown to be amongst one of the top traits by SFFS. It is the most potent feature in the SFFS ordering in contrast to its modest rating by MI.

5 Performance of Attractiveness

This section produces the analysis and experimental study into the capabilities of soft biometrics using attractiveness attributes for recognition. The identification of the unidentified subject was achieved by computing the sum of Euclidean distance between each subject in the probe set and all subjects in the gallery sets, resulting in a distance matrix. We can use the Euclidean distance to calculate the recognition accuracies with a Leave-one-out cross-validation (LOOCV) strategy by using a k-nearest neighbors (k-NN) classifier. Fig. 6 shows the accuracy by using all of the attributes in this study, which is 71.5 %. In comparison, a recognition rate of 57.21% is reported in [8] by using 24 comparative attributes. It is also noted that our dataset contains 100 subjects with four images per subject in total 400 subjects.

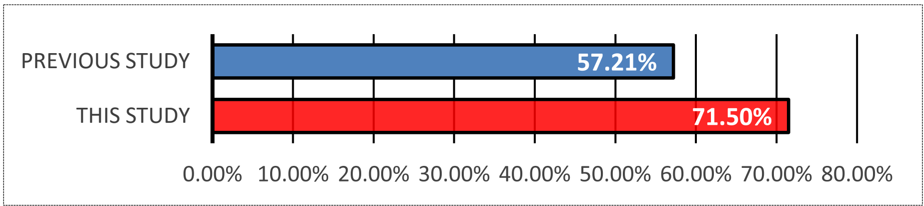

Fig. 6. Comparison the accuracy with previous study.

The Cumulative match characteristic (CMC) curve summarizes the identification accuracy, by the k-NN method, which scores the presence of the correct subjects. The CMC curves showed good identification performance using 22 soft facial traits in this study. In terms of verification, to measure the effect of the attractiveness. Using all attributes, recognition is still good (see Fig. 7) and the system performance increases once the attractiveness attribute is used as a trait in the recognition process.

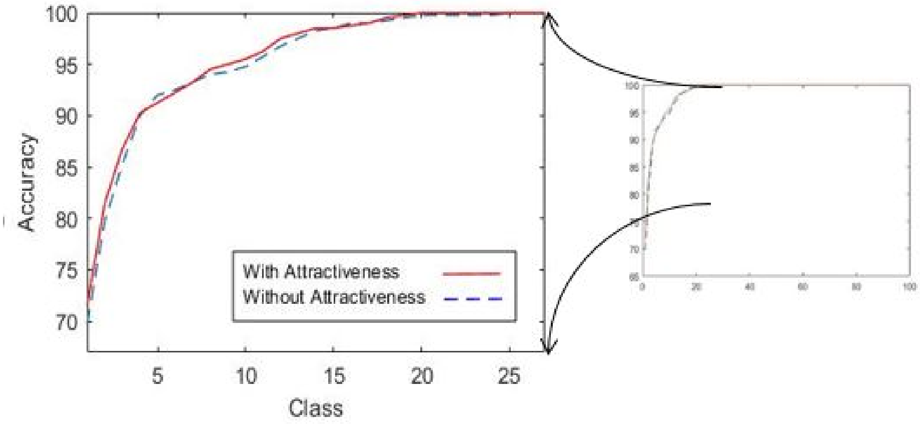

Fig. 7. Recognition via CMC performance on this study with and without attractiveness.

5.1 Abilities of Attractiveness

Fig. 8, shows the histograms of inter- and intra-class variations Manhattan distances with attractiveness, and without attractiveness. As shown in Fig. 8, attractiveness attribute improves the results since the proportion of false positives is reduced when attractiveness attribute is included with these study attributes.

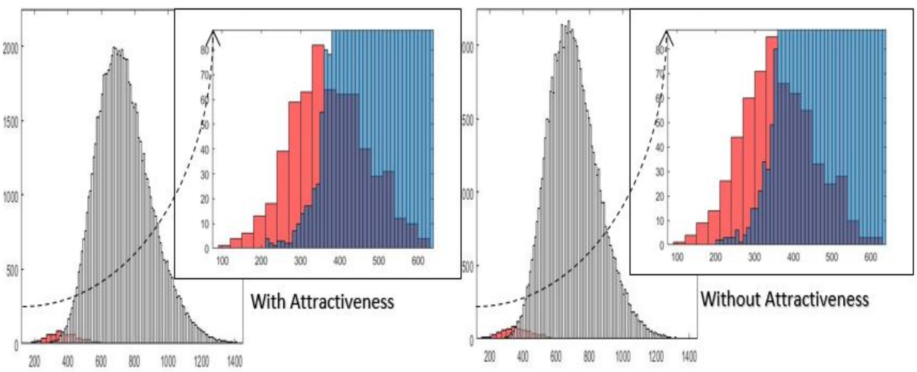

Fig. 8. Effect of attractiveness with all features of this study.

Attractiveness is used to complement the soft traits in subject recognition. Fig. 9 shows that the performance of the recognition system significantly improves when attractiveness is added to this study traits in comparison with the case where attractiveness has not been included in the set of the traits. We can establish the rate of accuracy with the attractiveness attribute in light of the investigation introduced in this section; the identification rate is improved when introducing attractiveness information or by up to 1.8% when introducing the whole set of soft biometrics attributes and the rate is improved 3.3 % when introducing gender information, as depicted in Fig. 9.

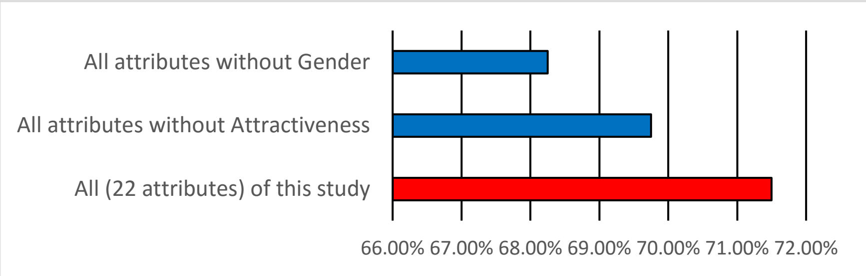

Fig. 9. Accuracy of 21 attributes with and without attractiveness, also with and without gender.

6 Conclusions and future work

This paper highlights the potential for facial attractiveness as a soft biometric tool. It presents an initial study of comparative facial attractiveness as a novel attribute for person description in a soft biometric framework for the purpose of recognition. To the best of our knowledge, this study is the first to use facial attractiveness in soft biometrics and the results of analyzing attractiveness through the designated attributes and suggests that there is advantage when facial attractiveness is used as a soft biometric.

One of our immediate future work is to use the whole LFW database. Humans can recognize the gender of a person, through some features. Our work on the correlations between attractiveness attributes can be used to improve gender recognition. It has not been known yet whether facial attractiveness is correlated to body attractiveness. This therefore gives an appropriate avenue for future research. This approach does not permit for collection of reliable features and also allows the estimation of the probability of facial attractiveness attributes from existing body attributes and vice versa. It will also be prudent to consider inclusion of automated approaches that assess attractiveness by computer vision. In this way, we intend to capitalize on and develop further this new approach.

References

1. Gonzalez-Sosa, E., Fierrez, J., Vera-Rodriguez, R. and Alonso-Fernandez, F., 2018. Facial Soft Biometrics for Recognition in the Wild: Recent Works, Annotation, and COTS Evaluation. IEEE TIFS, 13(8), pp.2001-2014

2. Zhang, H., Beveridge, J.R., Draper, B.A. and Phillips, P.J., 2015. On the effectiveness of soft biometrics for increasing face verification rates. CVIU, 137, pp.50-62.
3. Reid, D. Nixon, M. and Stevenage, S. (2013). Soft biometrics; human identification using comparative descriptions, IEEE TPAMI, 36(6), pp.1216-1228.
4. Tome, P. Fierrez, J. Vera-Rodriguez, R. and Nixon, M. (2014). Soft biometrics and their application in person recognition at a distance. IEEE TIFS, 9(3), pp.464-475.
5. Arigbabu, O., Ahmad, S., Adnan, W. and Yussof, S. (2014). Recent advances in facial soft biometrics. The Visual Computer, 31(5), pp.513-525.
6. Nixon, M. Correia, P. Nasrollahi, K. Moeslund, T. Hadid, A. and Tistarelli, M. (2015). On soft biometrics. Pattern Recognition Letters, 68, pp.218-230.
7. Parikh, D. and Grauman, K. (2011). Relative attributes. IEEE ICCV, pp.503-510.
8. Almudhahka, N., Nixon, M. and Hare, J. (2018). Semantic Face Signatures: Recognizing and Retrieving Faces by Verbal Descriptions. IEEE TIFS, 13(3), pp.706-716.
9. Samangooei, S. Guo, B. and Nixon, M. (2008). The use of semantic human description as a soft biometric. Proc. BTAS, pp.1-7.
10. Martinho-Corbishley, D., Nixon, M., & Carter, J. N. (2018). Super-Fine Attributes with Crowd Prototyping. IEEE TPAMI
11. Jaha, E. and Nixon, M. (2016). From clothing to identity: Manual and automatic soft biometrics. IEEE TIFS, 11(10), pp.2377-2390.
12. Shepherd, J. and Ellis, H. (1973). The Effect of Attractiveness on Recognition Memory for Faces. The American Journal of Psychology, 86(3), p.627.
13. Kagian, A. Dror, G. Leyvand, T. Cohen-Or, D. and Ruppin, E. (2007). A humanlike predictor of facial attractiveness, NIPS, pp.1-8.
14. Liu, S. Fan, Y. Samal, A. and Guo, Z. (2016). Advances in computational facial attractiveness methods. Multimed. Tools Appl. 75 (23) pp.16633–16663.
15. Chen, F. and Zhang, D. (2010). A benchmark for geometric facial beauty study. LNCS 6165, pp.21-32.
16. Thornhill, R. and Gangestad, S. (1999). Facial attractiveness, Trends Cognit. Sci., 3(12), pp.452-460.
17. Cunningham, M. (1986). Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty. Journal of Personality and Social Psychology, 50(5), pp.925-935.
18. Eisenthal, Y. Dror, G. and Ruppin, E. (2006). Facial attractiveness: Beauty and the machine, Neural Computation, 18, pp.119-142.
19. Xie, D. Liang, L. Jin, L. Xu, J. and Li, M. (2015). SCUT-FBP: A Benchmark Dataset for Facial Beauty Perception. Proc. IEEE SMC, pp.1821-1826.
20. Rhodes, G. (2006). The Evolutionary Psychology of Facial Beauty. Annual Review of Psychology. 57, pp.199-226.
21. Huang, G. Ramesh, M. Berg, T. and Learned-Miller, E. (2007). Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Tech. Report 07-49, Univ. Massachusetts, Amherst
22. Welinder, P. and Perona, P. (2010). Online crowdsourcing: rating annotators and obtaining cost-effective labels. Proc. IEEE CVPR Workshops, pp.25-32.
23. Guo, B. and Nixon, M. (2009). Gait Feature Subset Selection by Mutual Information. IEEE TSMC A, 39(1), pp.36-46.
24. Pudil, P., Novovičová, J. and Kittler, J., 1994. Floating search methods in feature selection. Pattern recognition letters, 15(11), pp.1119-1125.