

Automatic Semantic Face Recognition

Nawaf Yousef Almodhahka

Mark S. Nixon

Jonathon S. Hare

University of Southampton
Southampton, United Kingdom

{nya1g14,msn,jsh2}@ecs.soton.ac.uk

Abstract—Recent expansion in surveillance systems has motivated research in soft biometrics that enable the unconstrained recognition of human faces. Comparative soft biometrics show superior recognition performance than categorical soft biometrics and have been the focus of several studies which have highlighted their ability for recognition and retrieval in constrained and unconstrained environments. These studies, however, only addressed face recognition for retrieval using human generated attributes, posing a question about the feasibility of automatically generating comparative labels from facial images. In this paper, we propose an approach for the automatic comparative labelling of facial soft biometrics. Furthermore, we investigate unconstrained human face recognition using these comparative soft biometrics in a human labelled gallery (and vice versa). Using a subset from the LFW dataset, our experiments show the efficacy of the automatic generation of comparative facial labels, highlighting the potential extensibility of the approach to other face recognition scenarios and larger ranges of attributes.

I. INTRODUCTION

The adverse visual conditions of surveillance environments (e.g. low resolution, unconstrained pose and variable illumination), have increased the research interest in soft biometrics, which are physical and behavioural attributes that provide means of describing humans. The core idea is that these attributes can be used to retrieve identities where the acquisition of classical low level (hard) features (e.g. iris and fingerprint) become impossible [1]. The aim of semantic face recognition is to retrieve a suspect from a database of subjects using a human description of the suspect’s facial soft biometrics (e.g. an eyewitness statement) as illustrated in Figure 1. Soft biometrics can be effective with regard to human subjectivity, and provide invariance properties for describing visual appearance [2]. Thus, soft biometrics bridge the gap between machine and human vision [3]. Human face [10], body [17], and clothing [5], have been studied as sources of soft biometrics. Nonetheless, human faces have attracted the greatest interest as they are powerful for recognition at distance [1].

Work related to facial soft biometrics can be classified into two types: (1) categorical soft biometrics; in which each attribute is assigned to a particular class (e.g. the eyebrow is *thick*); and (2) comparative soft biometrics; in which an attribute is assigned for a subject relative to another subject (e.g. the *eyebrow* of subject *A* is *more thick* than that of subject *B*).

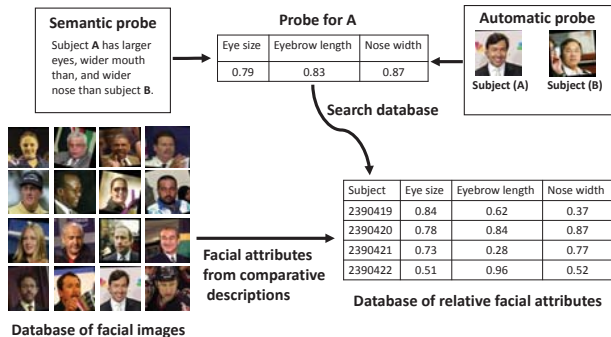


Fig. 1: Using comparative soft biometrics for people retrieval from a database of images.

Categorical facial soft biometrics have been explored for face verification [6], [7], [8] as well as recognition [9], [8]. All of these studies have proposed approaches for automatically extracting the soft biometrics from facial images. Comparative soft biometrics have shown a great impact on recognition accuracy [1], [10]. Human face recognition via comparative facial soft biometrics has been explored using both small constrained databases [11], [10] as well as larger unconstrained databases [2]. The studies on comparative soft biometrics, however, were focused on human generated comparative labels (e.g. through crowdsourcing) and did not address the automatic estimation of comparative labels from images or video footages, which is a key component towards the automatic retrieval and recognition of humans in surveillance scenarios. Moreover, the acquisition of comparative labels via crowdsourcing involves labour and time costs, which could be diminished if automatic estimation of comparative labels is possible.

This paper explores the problem of automatic estimation of comparative facial labels. We demonstrate our results using 400 images from the Labelled Faces in The Wild database (LFW) [12], which reflects challenging face recognition conditions such as variable pose, expression, position, illumination, and camera quality, in addition to occlusion of facial features. Furthermore, the paper aims to assess the impact of comparative labelling automation in different retrieval scenarios. To the best of our knowledge, the automatic estimation of comparative facial labels has never previously been addressed. Therefore, this study makes a major contribution to research on facial soft biometrics by demonstrating the

impact of automatically generated comparative labels in face recognition performance.

The main contributions of this paper are as follows:

- We explore the automatic estimation of comparative facial labels and assess their impact on unconstrained face recognition performance;
- We propose a framework for the automatic estimation of comparative facial labels from facial images which can be a baseline for the automatic estimation of comparative labels in general; and
- We present an evaluation for the significance of the automatically estimated labels and their correspondence with the human generated labels.

Section II describes the comparative facial soft biometrics used for this study and gives insights about the dataset used for the experiments. Section III explains the automatic estimation of labels with statistical views. Section IV illustrates the experimental design and presents the results with discussions. Lastly, Section V draws the conclusions and highlights the implications of the study.

II. COMPARATIVE FACIAL SOFT BIOMETRICS

The human face is considered as the most informative source for identification due to the significant discriminative power embedded within the facial components [1]. The significance and discriminative power of several comparative facial soft biometrics have been evaluated in [11], as well as [2], where the eyebrow and nose traits have shown high discriminative power. These findings agree to an extent with the study conducted in [13], which revealed that eyebrow is the most important component for human face recognition. Some of these attributes have previously been used in categorical and simile form [6], and in purely categorical form [9] and the eyebrows have a finer description here. One recent approach to attribute prediction [14] again predicts attributes many of which are similar to those here and also including face expression though the attributes are again categorical and were not used for recognition. One more recent study on relative attributes [15] concerned just noticeable differences and can determine whether the size of the eyes is distinguishable in two images, and that could be extended for these attributes. Equally, unsupervised learning could be used to analyse the features [16] and that is for the future. Since our objective is to study the automatic estimation of comparative labels for unconstrained human face recognition, which is usually coupled with adverse visual conditions of surveillance (e.g. variable illumination, pose, expressions, and resolution), we use the facial soft biometrics shown in Table I, in addition to the comparative labels acquired by [2] as they are collected via crowdsourcing for a subset of the LFW database, which emphasizes the unconstrained conditions of surveillance. Furthermore, the use of crowdsourcing for label acquisition, emphasizes the diversity of humans perception to the facial attributes, which better reflects the realistic conditions of face recognition [17]. Each of the attributes in Table I is associated with a binary (comparative) label that represents

TABLE I: Facial soft biometrics used in comparative labelling.

No.	Attribute	Comparative Labels
1	Chin height	[More small, More large]
2	Eye-brow hair colour	[More light, More dark]
3	Eye-brow length	[More short, More long]
4	Eye-brow shape	[More low, More raised]
5	Eye-brow thickness	[More thin, More thick]
6	Eye-to-eye-brow distance	[More small, More large]
7	Eye size	[More small, More large]
8	Face height	[More short, More long]
9	Face width	[More narrow, More wide]
10	Facial hair	[Less hair, More hair]
11	Forehead hair	[Less hair, More hair]
12	Inter eye-brow distance	[More small, More large]
13	Inter-pupil distance	[More small, More large]
14	Lips thickness	[More thin, More thick]
15	Mouth width	[More narrow, More wide]
16	Nose length	[More short, More long]
17	Nose septum	[More short, More long]
18	Nose-mouth distance	[More short, More long]
19	Nose width	[More narrow, More wide]
20	Spectacles	[Less covered, More covered]

TABLE II: Statistics on crowdsourcing of labels.

	Available	Possible	Coverage
Attribute comparisons	1381833	1596000	0.87
Subject comparisons	55065	79800	0.69
Comparisons per subject	275.32 (average)	399	-

the difference in the trait between two subjects as *Less* or *More* (e.g. Subject *A* has a *More Short* nose than Subject *B*). A comparative label is assigned with the integer -1 if it is of the class *Less*, the integer 1 if it is from the class *More*.

Dataset and annotations. The dataset used in [2] represents the training subset of View 1 in the LFW database, which consists of facial images for 4038 subjects. The comparative labels in [2] were collected via crowdsourcing resulting in 10065 subjects’ comparisons that were used to infer many more comparisons. For this study, we have randomly selected 400 subjects from the View 1 subjects, and extracted their pairwise comparisons from [2], where each comparison represents a sample comparative label. Also, it is important to mention that we have excluded the comparative label *Same*, since our approach emphasizes on relative rating based on differences among humans’ faces, not their similarities. Table II shows statistics about that dataset that is used in this paper. The term *available* in Table II refers to comparisons that were acquired via crowdsourcing and relation inference in [2], the term *possible* is defined as the maximum number of attainable comparisons between the subjects that is $n(n - 1)/2$ for a dataset of n items, and the term *coverage* refers to the ratio between *available* and *possible* comparisons that indicate the utilization achieved from the comparisons space, thus, the objective of the automatic estimation of comparative labels is to maximize the *coverage*. The facial images included in the dataset are all aligned using deep funnelling [18] and the inter-pupil distance for all the subjects is normalized to 50 pixels to ensure the consistency of comparisons between subjects.

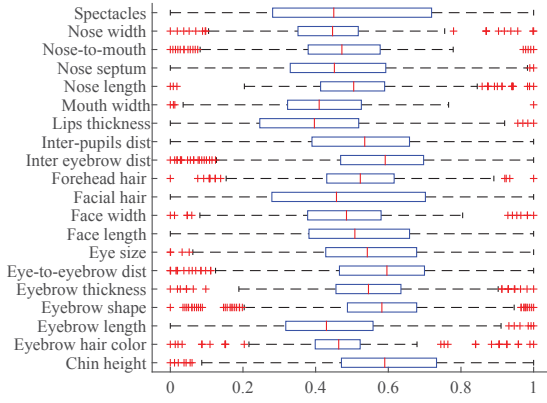


Fig. 2: Data distribution of the attributes.

Relative rating of attributes. Attribute rating is an essential step towards the creation of a biometric signature that can be used to uniquely identify a subject based on the relative strength of the subject’s attributes. The Elo rating system, which is a popular algorithm that is used for rating chess players, has been used to generate relative rates for facial soft biometrics based on the comparative labels in [10], [11], [2]. Accordingly, we use the Elo rating system in this paper for the same purpose. The main idea in the Elo rating system is to compute a player’s rate based on the difference between the player’s expected and actual game outcomes. Thus, for a game between two players A and B , their expected scores, E_A and E_B respectively, are calculated as follows:

$$E_A = \left[1 + 10^{(R_B - R_A)/400} \right]^{-1} \quad (1)$$

$$E_B = \left[1 + 10^{(R_A - R_B)/400} \right]^{-1} \quad (2)$$

where R_A and R_B are the current rates of players A and B respectively. Depending on the game outcome, the score, S , of a player is set as: 1, 0.5, or 0; for win, tie, or loss, respectively. Consequently, the updated rates, \bar{R}_A and \bar{R}_B , for players A and B correspondingly, become:

$$\bar{R}_A = R_A + K(S_A - E_A) \quad (3)$$

$$\bar{R}_B = R_B + K(S_B - E_B) \quad (4)$$

where K is the score adjustment parameter, which determines the sensitivity of rate update. In the context of our paper, players are the subjects in the dataset, and games are the comparisons between the subjects. For an attribute X , a comparison outcome is mapped to a subject score, S , as: 1, 0.5, or 0, for the labels *More X*, *Same X*, or *Less X* respectively. As explained earlier, the *Same X* label is not considered in this paper as our approach emphasizes on differences between people. Thus, the *tie* situation in the Elo rating system is not applicable for our approach.

Attribute analysis. The attribute analysis is carried out using the relative rates generated for each attribute by the Elo rating system explained earlier in this section. It can be inferred

from the boxplot shown in Figure 2 that the attributes do not follow a normal distribution. In addition, most of the attributes’ distributions are skewed and have outliers. Altogether, the boxplot reveals that the dataset is challenging. Also, it is important to mention that the outliers are not eliminated in the analysis and experiments performed throughout this paper, as these data points represent attributes of subjects in the dataset. Furthermore, the discriminative power of the attributes is assessed using mutual information [19], which reveals the amount of information contained in each attribute about subjects [11], and the results are shown in Figure 3, where some interesting findings emerge. Firstly, *eyebrow shape* and the distance attributes of the eye-eyebrow region (i.e. *inter-pupil distance*, *inter-eyebrow distance*, and *eye-to-eyebrow distance*) have high discriminative power, which support the emphasis on this region of the human face [11]. Secondly, *facial hair*, has high discriminative power that might be attributed to its binary nature and association with gender. Thirdly, some eyebrow attributes have, in general, lower discriminative potential, which can be referred to the outliers resulted from the exposure of eyebrows to beauty interventions. Fourthly, *forehead hair* has the lowest discriminative power, probably due to its high variance. Finally, the correlation analysis based on Pearson’s r (as shown in Figure 4) reveals that the correlation between the attributes are either small or insignificant, which indicate the independence of the attributes used in this paper.

III. APPROACH

The research in face recognition using comparative soft biometrics has, to date, tended to focus on exploring retrieval using human generated annotations in a human annotated gallery [10], [11], [2]. One key question that needs to be asked, however, is whether comparative facial labels can be estimated automatically for a database of facial images. In this paper, we address this question by investigating retrieval using human labelled facial attributes in a gallery of automatically labelled attributes. We present a baseline approach for estimating comparative facial

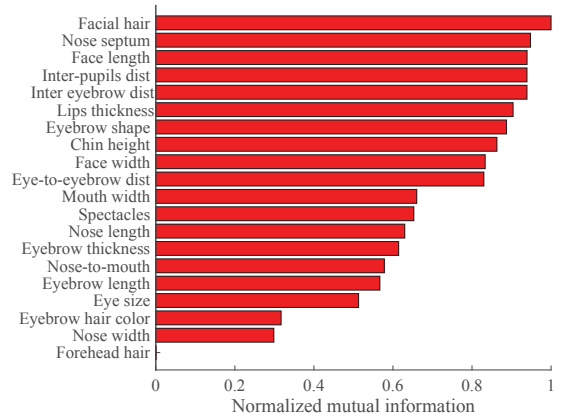


Fig. 3: Discriminative power of attributes using mutual information.

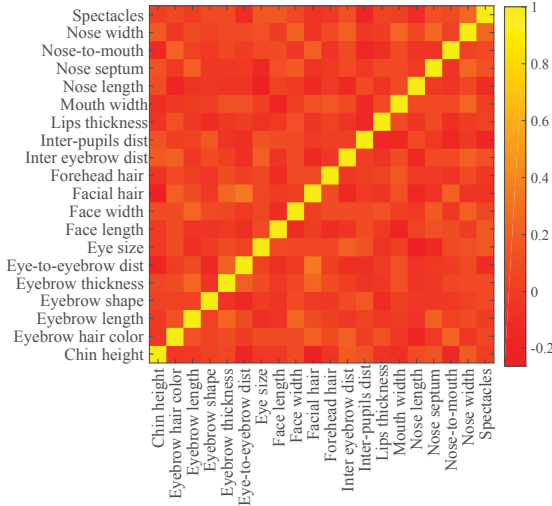


Fig. 4: Map of correlations between attributes.

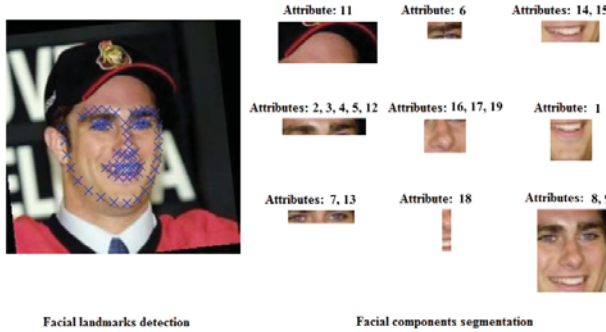


Fig. 5: Facial landmarks detection with the DPM [20] and facial components segmentation along with the corresponding attributes. Visual features for each attribute are extracted using GIST [21] descriptor from the images of facial components. The attributes are numbered as in Table I.

labels automatically and producing biometric signatures, which consist of the relative rates of the attributes. Our approach has two main stages: (1) facial feature detection and extraction; and (2) estimation of comparative labels.

Facial features detection and extraction. The aim of this stage is to segment the facial components that correspond to the attributes list in Table I, and producing the corresponding visual biometric signatures. To detect the facial landmarks, we use a deformable part model (DPM) [20], which is trained for face alignment in the wild, for locating 66 facial landmarks, then the major facial components are segmented as shown in Figure 5. After segmenting the major facial components, a visual descriptor vector that corresponds to each facial component is generated using GIST features [21]. The procedure of extracting GIST features starts by performing intensity normalization for the facial component image, then passing it through a series of Gabor filters in 4 scales and 8 orientations per scale. The resulted 32 orientation maps are divided into 4×4 grids, and

the mean intensity is calculated for each block in the grid to form a feature. This results in a vector of 512 features for each facial component. These visual features are used to train a model to estimate comparative facial labels as explained next in this section.

Estimation of comparative labels. We have mentioned in Section II that each attribute is associated with a binary comparative label that represents the difference between a pair of subjects as *Less* or *More*, which are mapped to -1 and 1 respectively. To automatically estimate comparative labels from the images set in this paper, we train 20 multiple linear regressors, in which the visual features are considered as independent variables that are used to predict a dependent variable (i.e. a comparative label). Linear regression is chosen to estimate comparative labels as it explicitly includes order. The objective of training the regressors is to model a linear relationship between visual features, which are extracted using the GIST descriptor, and comparative label values (-1 or 1) for the 20 facial attributes (see Table I). The dataset is randomly divided into 10 folds, from which 9 folds that have a total of m comparisons are used to train the regressors, while the remaining fold (consisting of n comparisons) is used for testing, such that the total number of comparisons for an attribute is $m + n$. The relationship between the visual features of an attribute $1 \leq t \leq 20$ and the associated comparative labels can be modelled as:

$$y_c^t = \beta_0 + \beta_1 x_{c1}^t + \beta_2 x_{c2}^t + \dots + \beta_p x_{cp}^t + \epsilon_t \quad (5)$$

where $1 \leq c \leq m$ is the comparison index, y_c^t is the value of the comparative label $y \in \{-1, 1\}$ of the comparison c for the attribute t , x_{ck}^t is the difference in the k^{th} visual feature between the two subjects forming the comparison c for the trait t , while $1 \leq k \leq p$ is the visual feature index, and ϵ_t is the error term. The objective of training is to estimate a vector of weight coefficients $\hat{\beta} = \{\beta_0, \beta_1, \dots, \beta_p\}$ such that the sum of squared residuals is minimized:

$$\text{minimize} \left(\frac{1}{m} \sum_{c=1}^m (y_c^t - \hat{y}_c^t)^2 \right) \quad (6)$$

where y_c^t is the value of the comparative label to be predicted, while \hat{y}_c^t is the predicted comparative label. After the weight coefficients vector, $\hat{\beta}$, is estimated, the predicted comparative label b for an attribute t between subjects (i, j) forming a comparison in the testing fold is derived as follows:

$$b_d^t(i, j) = \begin{cases} -1 & \text{if } \hat{y}_d^t < 0 \\ 1 & \text{if } \hat{y}_d^t \geq 0 \end{cases} \quad (7)$$

where $1 \leq d \leq n$ is the index of the comparison in the testing fold. The predicted comparative labels are used to infer relative rates for the 20 attributes using Elo rating as described in Section II. The relative rates of the 20 facial soft biometrics are used to construct a biometric signature for each subject, which consists of 20 relative rates representing the facial soft biometrics, and it is consequently used for subject retrieval as explained next in Section IV.

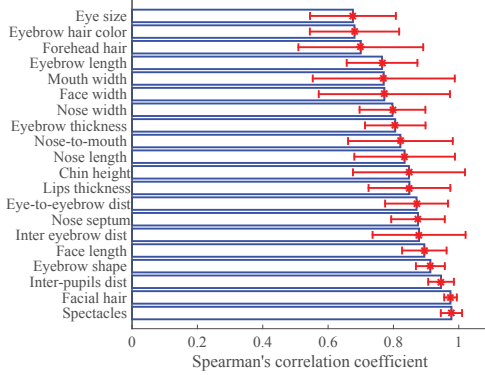


Fig. 6: Correspondence between human generated and machine generated labels averaged over 10 folds and measured with Spearman’s correlation coefficient.

Figure 6 shows the correspondence between human generated and machine generated comparative facial soft biometrics measured using Spearman’s correlation coefficient between the relative rates derived from the human generated labels and the relative rates derived from machine generated labels averaged over 10 folds. From Figure 6, it can be seen that there is a notable association between the discriminative power analysis (Figure 3) and the correspondence analysis. The binary-like attributes (i.e. *spectacles* and *facial hair*) have the strongest correspondence, followed by the distance attributes for the eye-eyebrow region. While *eye size* and *eyebrow hair colour* have the lowest, yet significant, correspondence. These results show that the automatic estimation of comparative facial labels is significantly dependent on the discriminative power of the facial attributes, which highlights the importance of choosing soft biometrics that are clear and self-descriptive for human annotators.

IV. FACE RECOGNITION FOR RETRIEVAL

The objective of semantic face recognition is to retrieve a suspect from a database of subjects using a biometric signature that is generated from the relative measurements of the suspect’s facial soft biometrics. The experiments conducted in this paper simulate an operational scenario in which a database of facial images (gallery) is searched using an eyewitness statement (i.e. semantic descriptions). As explained in Section II, the dataset used in this paper consists of facial images for 400 subjects from the LFW database that were labelled comparatively using crowdsourcing. The performance evaluation in this experiment is based on 10-fold cross validation. The test runs through 10 iterations covering each fold. In each test iteration, one of the folds is used as a testing set, while the remaining 9 folds are used to train a set of 20 multi-linear regressors to estimate comparative labels for the 20 facial attributes. The experiment has two steps: gallery construction and subject retrieval.

A. Gallery Construction

Two types of galleries are created and evaluated in our experiments: *human* and *automatically* labelled. The *human gallery* is constructed by removing (withholding) pairwise

comparisons between each probe and 10 randomly selected subjects from the 9 training folds, and using the remaining comparisons to generate a biometric signature for each subject in the gallery. The 10 *withheld* subjects are used to generate a probe biometric signature for each subject for retrieval by comparing them relative to the probe (as will come next in this section). The number of these counterpart subjects is set to 10 as it is the average size of an identity parade [22]. The *automatic gallery* is constructed from automatically estimated comparative labels for each subject in the testing fold from the visual features using the trained regressors as described in Section III, and a biometric signature is generated for each subject in the dataset. The comparisons (i.e. facial image pairs) that are used to train the regressors are different from those used for retrieval.

B. Subject Retrieval

The probes used for retrieval in our experiments are classified into: (1) *human probe*, for which a biometric signature that is composed of relative rates for the 20 attributes is generated using the Elo rating from human comparisons (i.e. crowdsourced) between the probe and the 10 subjects that were *withheld* during gallery construction, and this probe can be considered as an eyewitness statement; and (2) *automatic probe*, for which biometric signature is formed from automatically estimated comparative labels generated from the probe’s facial image and the facial images of the same 10 subjects who were randomly selected to construct the *human probe*. The similarity between the probe and each subject in the gallery is measured using the Pearson’s distance, d_P , as follows:

$$d_P = 1 - \frac{\sum_{t=1}^{20} (X(t) - \bar{X})(Y(t) - \bar{Y})}{\sqrt{\sum_{t=1}^{20} (X(t) - \bar{X})^2} \sqrt{\sum_{t=1}^{20} (Y(t) - \bar{Y})^2}} \quad (8)$$

where X is the biometric signature of the probe, Y is the biometric signature of the subject in the gallery being compared with the probe, and $1 \leq t \leq 20$ is the trait index. The subject in the gallery that results in the minimum distance with the probe (i.e. the nearest neighbour) is considered as a match. This procedure is iterated over all the subjects of the test fold, running over the 10 folds, and the experiment is repeated 30 times. The harmonic mean of returned ranks for the 30 trials is used to calculate the retrieval accuracies for each of the 10 folds, and the arithmetic mean of the 10 folds is considered as the experiment outcome. We explore three different retrieval scenarios based on this experimental procedure: *human probe* with *automatic gallery*; *human probe* with *human gallery*; and *automatic probe* with *human gallery*. The results of retrieval using these scenarios are reported using Cumulative Match Curve (CMC) as shown in Figure 7

As can be seen from Figure 7, retrieval using *human probe* in the *automatic gallery* slightly outperforms retrieval in the *human gallery* at rank-1 achieving an accuracy of 83.75%. Starting from rank-2, retrieval using *human probe* in the *human gallery* shows better accuracy as compared to the

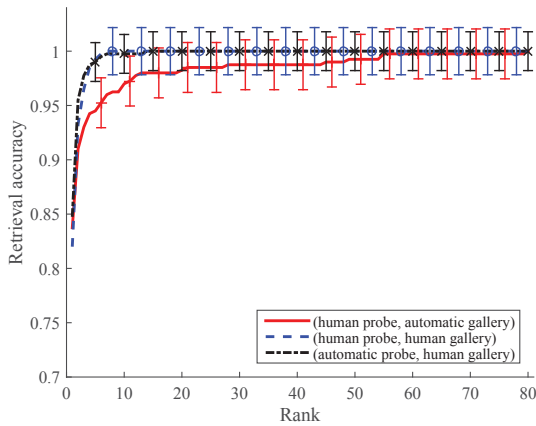


Fig. 7: Retrieval accuracy using 10 subject comparisons.

automatic gallery. Yet, retrieval using *human probe* in the *automatic gallery* achieves a rank-10 accuracy of 97.25%, and reaches 100% accuracy at rank-80, which implies that a match will be always found in the top 20% results returned from the retrieval. Also, retrieval using *automatic probe* in the *human gallery* yields an almost similar accuracy to retrieval using *human probe* in the *human gallery*, which reveals the effectiveness of our approach for automatically estimating a biometric signature from suspect and subjects facial images. In [9], 46 categorical facial attributes were defined, and identification was explored with 1196 subjects from the FERET database achieving rank-1 retrieval rate of 19% using *human probe* in *automatic gallery*, and 22.5% using *human probe* in *human gallery*. In [6], 65 binary attributes, which recognize the presence or absence of a visual trait, are used with the LFW database for verification, and they achieved an accuracy of 83.62% with the entire LFW dataset (13233 images for 5749 subjects). Comparing our work with [6], it is important to emphasize that our objective is to explore the automatic estimation of comparative labels, and evaluate retrieval using semantic descriptors in an automatically labelled database. Therefore, we use only a subset of the LFW database. Moreover, we use fewer attributes (i.e. the 20 facial attributes proposed by [2]), as it will be more practical for eyewitnesses to recall 20 attributes, compared to 65 attributes proposed in [6].

Overall, the results of our experiments indicate that the automatic estimation of comparative facial labels can result in significant improvements with respect to retrieval accuracy. Furthermore, the results show that the incorporation of machine learning with computer vision techniques can provide efficient approaches for the comparative labelling of facial attributes, and hence, can reduce the involvement of human annotators for labelling facial images in a database.

V. CONCLUSIONS

This paper explores the automation of face recognition using comparative soft biometrics in addition to assessing the impact of automatic estimations of comparative labels on face retrieval performance. We have proposed an approach that incorporates computer vision with machine learning

for automatically estimating comparative facial labels. In addition, we have examined a scenario in which a database of facial images is automatically labelled, then searched using a semantic description for the subject to be retrieved. Using a subset of the LFW database, we have shown that automatic labelling of comparative facial soft biometrics can result in significant retrieval accuracy while reducing dependency on human annotators. These findings extend the horizons of bridging the semantic gap between humans and machines, in addition to serving as a base for future studies on automatic labelling of comparative soft biometrics.

REFERENCES

- [1] Daniel Reid, Sina Samangooei, Cunjian Chen, Mark Nixon, and Arun Ross. Soft biometrics for surveillance: an overview. *Machine learning: theory and applications*. Elsevier, pages 327–352, 2013.
- [2] Nawaf Almadhahka, Mark Nixon, and Jonathon Hare. Unconstrained human identification using comparative facial soft biometrics. In *IEEE BTAS*, 2016.
- [3] Mark Nixon, Paulo Correia, Kamal Nasrollahi, Thomas Moeslund, Abdenour Hadid, and Massimo Tistarelli. On soft biometrics. *Pattern Recognition Letters*, 68:218–230, 2015.
- [4] Daniel Martinho-Corbishley, Mark Nixon, and John Carter. Soft biometric retrieval to describe and identify surveillance images. In *IEEE ISBA*, 2016.
- [5] Emad Jaha and Mark Nixon. Soft biometrics for subject identification using clothing attributes. In *IEEE IJCB*, 2014.
- [6] Neeraj Kumar, Alexander Berg, Peter Belhumeur, and Shree Nayar. Attribute and simile classifiers for face verification. In *IEEE ICCV*, 2009.
- [7] Neeraj Kumar, Alexander Berg, Peter Belhumeur, and Shree Nayar. Describable visual attributes for face verification and image search. *IEEE TPAMI*, 33(10):1962–1977, 2011.
- [8] Pedro Tome, Ruben Vera-Rodriguez, Julian Fierrez, and Javier Ortega-Garcia. Facial soft biometric features for forensic face recognition. *Forensic science international*, 257:271–284, 2015.
- [9] Brendan Klare, Scott Klum, Joshua Klontz, Emma Taborsky, Tayfun Akgul, and Anil Jain. Suspect identification based on descriptive facial attributes. In *IEEE IJCB*, 2014.
- [10] Daniel Reid and Mark Nixon. Human identification using facial comparative descriptions. In *IEEE ICB*, 2013.
- [11] Nawaf Almadhahka, Mark Nixon, and Jonathon Hare. Human face identification via comparative soft biometrics. In *IEEE ISBA*, 2016.
- [12] Gary B Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, Technical Report 07-49, University of Massachusetts, Amherst, 2007.
- [13] Javid Sadr, Izzat Jarudi, and Pawan Sinha. The role of eyebrows in face recognition. *Perception*, 32(3):285–293, 2003.
- [14] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *IEEE ICCV*, 2015.
- [15] Aron Yu and Kristen Grauman. Just noticeable differences in visual attributes. In *IEEE ICCV*, 2015.
- [16] Chen Huang, Chen Change Loy, and Xiaoou Tang. Unsupervised learning of discriminative attributes and visual representations. In *IEEE ICCV*, pages 5175–5184, 2016.
- [17] Daniel Martinho-Corbishley, Mark Nixon, and John Carter. Soft biometric recognition from comparative crowdsourced annotations. In *ICDP*, pages 1–6. IET, 2015.
- [18] Gary Huang, Marwan Mattar, Honglak Lee, and Erik G Learned-Miller. Learning to align from scratch. In *NIPS*, 2012.
- [19] Baofeng Guo and Mark S Nixon. Gait feature subset selection by mutual information. *IEEE SMC*, 39(1):36–46, 2009.
- [20] Georgios Tzimiropoulos and Maja Pantic. Gauss-newton deformable part models for face alignment in-the-wild. In *ICPR*, 2014.
- [21] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 2001.
- [22] Daniel A Reid, Mark S Nixon, and Sarah V Stevenage. Soft biometrics; human identification using comparative descriptions. *IEEE TPAMI*, 36(6):1216–1228, 2014.