Soft Biometrics; Human Identification using Comparative Descriptions

Daniel A. Reid, Mark S. Nixon, Sarah V. Stevenage

Abstract—Soft biometrics are a new form of biometric identification which use physical or behavioral traits that can be naturally described by humans. Unlike other biometric approaches, this allows identification based solely on verbal descriptions, bridging the semantic gap between biometrics and human description. To permit soft biometric identification the description must be accurate, yet conventional human descriptions comprising of absolute labels and estimations are often unreliable. A novel method of obtaining human descriptions will be introduced which utilizes comparative categorical labels to describe differences between subjects. This innovative approach has been shown to address many problems associated with absolute categorical labels - most critically, the descriptions contain more objective information and have increased discriminatory capabilities. Relative measurements of the subjects' traits can be inferred from comparative human descriptions using the Elo rating system. The resulting soft biometric signatures have been demonstrated to be robust and allow accurate recognition of subjects. Relative measurements can also be obtained from other forms of human representation. This is demonstrated using a support vector machine to determine relative measurements from gait biometric signatures - allowing retrieval of subjects from video footage by using human comparisons, bridging the semantic gap.

Index Terms—Soft Biometrics, Human Descriptions, Retrieval, Comparisons, Regression, Gait Biometrics

✦

1 INTRODUCTION

Traditional biometric techniques identify people using distinct physical or behavioral features. These features are clearly discriminative although they can rarely be described using linguistic labels. This is known as a semantic gap. This restricts identification to situations in which the subject's biometric signature can be obtained and only permits identification of those subjects whose biometric signature has previously been recorded. Soft biometrics are a new form of biometric identification which concerns labels that people use to describe each other, like 'tall', 'skinny' and 'male'. Although each label can have reduced discriminative capability, they can be combined for identification [1], [2] and fusion with traditional 'hard' biometrics [3], [4]. Dantcheva et al. [5] likens this to obtaining a single ridge of a fingerprint or a small section of the iris. These would not be sufficient to identify a subject but by gathering many small features we are able to build a unique biometric signature.

One of the main advantages of soft biometrics is their relationship with human description: humans naturally use soft biometric traits to identify and describe each other. This permits identification and retrieval based solely on a human description of the subject, possibly obtained from an eyewitness.
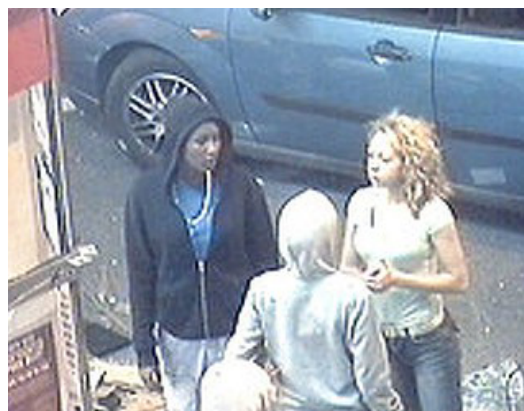

Fig. 1. Surveillance frame displaying common surveillance problems[1]

Though face and gait are the only practical biometrics at a distance, in surveillance scenarios they can suffer from low frame rate and/or resolution. Figure 1 shows an example of a typical CCTV video frame showing looters at the 2011 London riots. It can be observed that although the picture is at low resolution a detailed human description of the subjects can still be given. In comparison, automatic facial recognition would struggle with the low resolution and non-frontal viewpoint. Soft biometric traits can be obtained from the data derived from low quality sensors, including surveillance cameras. Soft biometrics also require less computation compared to 'hard' biometrics, no cooperation from the subject and are non-invasive - making them ideal in surveillance applications.

- *D.Reid and M.Nixon are with the School of Electronics and Computer Science, University of Southampton, Southampton, UK. S.Stevenage is with the School of Psychology, University of Southampton, UK*

1. Metropolitan Police Flickr Account

To allow identification from human descriptions, physical properties must be accurately described. Conventional human descriptions represent an important element of eyewitness evidence, although they can be considered inaccurate and unreliable [6], [7]. They consist of either absolute categorical labels (e.g. 'tall') or estimates of human characteristics (e.g. 190cm). Previously, categorical labels have been used in a soft biometric system [1]. Given that humans can be inaccurate when predicting measurements [8], labels were seen as a more robust method of obtaining human descriptions. One major problem associated with absolute categorical labels is their highly subjective nature. A label's meaning is based on the person's own attributes and their own perception of population averages and variation. This can vary, making absolute labels less reliable. Categorical labels naturally lack detail, resulting in biometric signatures which have poor discriminatory capability. This paper introduces a new method for obtaining human descriptions which exploits the process of making visual comparisons between subjects.

Comparing the appearance of two subjects is a very natural process and may be more reliable than the use of absolute labels because comparisons are assigned based on a specified benchmark resulting in a more objective description. We exploit the ease of making comparisons to explore a new method to provide reliable and robust descriptions.

A set of *relative measurements* describing the subject can be accurately inferred from comparative labels (that are derived from comparing one person with another, e.g. 'taller') and used as a biometric signature. These signatures have been shown to be highly discriminative allowing accurate biometric retrieval.

The novelty of this paper is to demonstrate how individuals can be identified from a database of videos based solely on a verbal comparative description, with particular contribution of

- Extended analysis of the benefits of comparative labels over absolute
- Extended description of the derivation of relative measures from comparative labels
- Demonstration of capability to learn soft biometric labels and relative measurements from video
- Demonstration of capability to retrieve subjects from a database by verbal description

The remainder of this paper will explore the effectiveness of human comparisons and how they can be applied to soft biometric recognition and retrieval. Section 2 will investigate related work in psychology, biometrics and computer vision. An introduction to human comparisons and the database used throughout this paper will be presented in section 3. Finally, section 4 will explore how to utilize comparisons for identification. This will include an introduction to the Elo ranking system which is used to infer relative measurements and the identification of subjects from soft biometric databases and video footage.

2 RELATED WORK

2.1 Psychology of Human Descriptions

To allow identification from human descriptions, the physical properties described must be accurate, salient and reliable. Human descriptions generally consist of two types of descriptions: labels and measurements. Labels are predominantly used to describe inherently categorical traits like ethnicity and gender, but they can also be used to describe continuous traits, for instance a height description can include 'short', 'medium' and 'tall'. Estimations of continuous traits are more commonly described using measurements detailing the feature's length, width or weight. Much research has been conducted into obtaining accurate human descriptions due to their importance in many criminal investigations.

Kuehn studied the descriptions provided by victims in 100 police investigations [9]. Despite the prominence given to faces in eyewitness testimony, it was discovered that gender, age, height, build, race, weight, complexion, and hair color were mentioned over 70% of the time, whilst facial features were rarely mentioned. Similarity, MacLeod et al. investigated the prominence and reliability of whole body descriptors [10]. 13 of the most reliable body descriptors were identified from a total of 1238, the most reliable being thin-fat, short-tall, thin-thick legs, slim-barrel chest and short-long legs.

Yuille and Cutshall [8] showed that estimates of height, weight and age were incorrect 50% of the time based on 95 cases (considered accurate if within 2 inches, 5 pounds and 2 years respectively of the actual measurement). Inaccurate estimates have been attributed to an own anchor effect, where the witness's own characteristics were used as a point from which to judge the suspect [11]. It was also found that descriptions tended to show a regression to the mean, or to what the witness estimated as the mean. Shorter people were estimated as taller than they really were, and vice versa. This was thought to occur due to the witness shying away from extreme judgments [7].

The current paper introduces a new form of human description for soft biometrics which was systematically designed to avoid the limitations of absolute labels and continuous estimations whilst building on the analysis of the content of human descriptions.

2.2 Human Descriptions and Soft Biometrics

Soft biometrics can be split into three distinct categories: bolstering of hard biometrics utilizing measured soft biometric traits, identification using measured soft biometric traits and identification using verbal descriptions of soft biometric traits. This paper
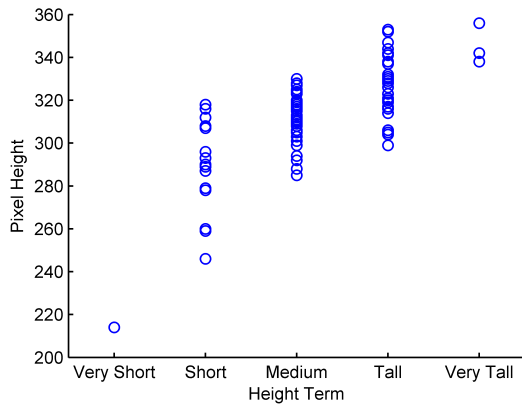
Fig. 2. The relationship between pixel height and absolute labels

Fig. 3. Retrieval accuracy of absolute descriptions from a soft biometric database

presents an identification technique which utilizes verbal descriptions of soft biometric traits and as such this section will examine how verbal descriptions have previously been used in soft biometrics. For a survey of soft biometrics as a whole please refer to [12].

There are two distinct methods for obtaining information about soft biometric features. The first automatically identifies features from data describing the subject, most commonly in the form of imagery. The second aims to generate biometric information from verbal descriptions of the subject.

Samangooei and Nixon [1] developed a soft biometric system which identifies subjects from video footage (Soton gait database [13]) based solely on a verbal human description. This description was composed of 23 absolute categorical labels which were chosen to be universal, distinct, easily discernible at a distance and largely permanent. The selected soft biometric traits featured both intrinsically categorical attributes, like hair color, and characteristics generally associated with value metrics, like height - both were described using absolute labels.

Initially 959 descriptions of the 115 subjects from the Soton gait database were obtained and used to build a database of soft biometric feature vectors which described the given descriptions. Initial analysis of the descriptions showed that the categorical labels used to describe the subjects were unreliable, especially when describing traits generally associated with value metrics. Figure 2 shows the relationship between the height of the subjects (obtained from the video footage and represented in pixels) and the median absolute height label used to describe the subjects. Large overlaps between the short, medium and tall labels were observed resulting in a statistically significant ($p < 0.0001$) Pearson's correlation of 0.71. This incorrectness between actual and labeled height is due to the categorical and highly subjective nature of the labels.
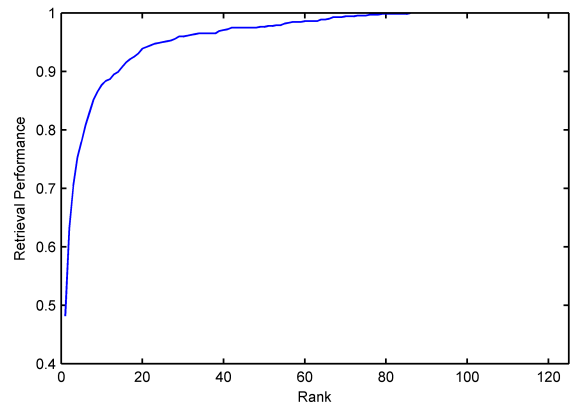
The database of soft biometric feature vectors was used by the authors to assess the discriminatory power of the descriptions. Each subject's feature vector consisted of the most commonly used label to describe a subject's soft biometric trait (on average each subject was described by 8 individual annotators). Recognition experiments were conducted by retrieving subjects from the database to assess the uniqueness of each subject's soft biometric feature vector and the variance between multiple descriptions of the same subject. A leave-one-out validation approach was used to evaluate the recognition performance. The probe, which was used to query the database, was formed from a single verbal description of the subject given by a single annotator. The remaining descriptions of the subject were averaged and used as the gallery, the feature vector within the database being searched. Figure 3 shows the results. The rank 1 retrieval performance (i.e. the recognition accuracy) was found to be 48%. Retrieval performance increased to 90% at rank 15. Subject interference [5] is a known problem when using labels and occurs when two subjects are indistinguishable from each other due to the limited number of labels available. This obviously has a drastic effect when attempting to identify a subject and would explain the poor recognition results. This highlights the lack of distinctiveness between subjects due to the limited information conveyed using categorical labels. As such, absolute labels can be used to recognize people but are limited in accuracy leading to a limited recognition capability.

Denman et al. [14] used soft biometric traits to identify people using previous observations or human descriptions when traditional biometrics are unavailable. The height and colour of the torso, legs, and head are used to model subjects. Identifying these three body components is done by first locating the person using background segmentation and then analysing the colour of moving pixels in each row. Large colour differences can often be found between the head, torso

and legs due to clothing that can be easily identified by examining colour gradients. The PETS 2006 surveillance database was used to test the system. This dataset features four cameras monitoring a train station: four recordings of 25 people were obtained. The system achieved an equal error rate of 6.1% when identifying individuals from previous observations. Unfortunately the performance of identification utilizing verbal descriptions was not detailed.

2.3 Relative Information

Relative information has recently been explored to improve human descriptions of objects within images. Several techniques have exploited similarities between objects as a form of description. Kumar et al. [15] have explored similarities between faces to identify and explain facial attributes. The developed 'simile classifiers' recognize similarities between a face (or regions of a face) and a set of specific reference subjects. This allows descriptions such as '*lips like Barack Obama*' or '*a nose like Owen Wilson*'. The advantage of this system is the ability to produce descriptions of features which are generally hard to describe. Wang et al. [16] exploits similarities between objects to allow recognition with few or no examples. Descriptions such as '*a zebra is similar to a horse in shape and a crosswalk in texture*', allows the approach to identify a zebra with no training examples. Exploiting descriptions of similarity between objects has been shown to improve recognition of objects within images with few training examples. Both of these techniques utilize relative information to improve descriptions, although they differ significantly from our approach. Similarity between reference subjects or other objects provides a method of description, whereas the comparing of subjects provides an ordering based on the specific trait being compared. Although different, these techniques show the benefits of relative information especially when describing features or attributes which are normally difficult to communicate.

Image descriptions have been further improved by determining order based on the strength of a specific attribute, allowing such comparisons as '*lions are larger than dogs*' [17]. Given a set of images and a partial set of comparisons detailing the relative strength of a certain attribute, the technique determines a complete ordering of the images. This was approached as an optimization problem where the comparisons were treated as constraints. A ranking support vector machine was used to determine a ranking function which fitted a weight vector to maximize the number of constraints satisfied - this was based on ranking algorithms used within search engines [18]. The ranking function could then be used to determine the ordering between all of the images. Zero-shot learning from relationships was introduced based on this ordering approach, allowing previously unseen objects

to be identified based on comparisons with observed objects. The zero-shot learning results show that the relative descriptions convey stronger discriminatory power compared to binary descriptions.

3 HUMAN COMPARISONS FOR SOFT BIO-METRICS

In this section we detail our new approach that uses comparisons to address the difficulties associated with conventional human descriptions.

3.1 Human Comparison Database

The method used to obtain descriptions from an observer is an important consideration when exploring a new form of human description. In the case of human comparisons the practical limitations of human memory and the ability of humans to compare bodily attributes must be considered and explored. An experiment was designed to answer the following questions:

- Do relative measurements provide more discriminatory information than absolute labels?
- Are the resulting relative measurements highly correlated with the subject's physical attributes?
- Is the developed method of obtaining human comparisons practical?

Although descriptive, a single comparison between a suspect and another person will only explain the differences between the two. Thus, the inferred physical traits of the suspect will depend on the subject they were compared to. Multiple comparisons must be available to infer a more robust description, with each comparison allowing the description of the suspect to be refined. A practical method of obtaining comparisons, between a *target* subject (representing the suspect in application settings) and multiple *subjects*, is to present videos of the subjects to the annotator. This permits multiple comparisons with minimal equipment and personnel. To validate this approach the experiment will present videos of subjects from the Soton gait database [13] to the annotator. The gait database includes videos of 100 people walking in a plane normal to the view of the camera. Previously absolute categorical labels had been collected for the same database [1] - allowing comparisons to be drawn between the two forms of description.

Comparisons were gathered using the website shown in figure 4. The website was designed to allow videos of both the subject and target to be presented to the annotator simultaneously. This allows users to make direct comparisons without memory demands or uncertainties concerning the scale of the videos. Drop-down boxes for each trait allowed users to describe how the subject differed from the target. The chosen label was emphasized by constructing a sentence explaining the given annotation - ensuring
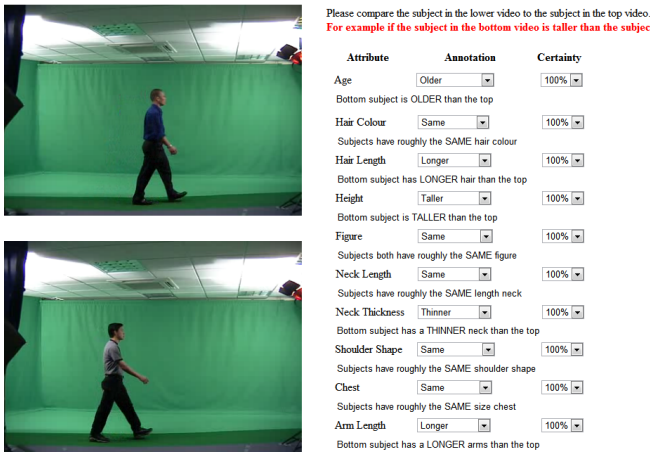
Fig. 4. Comparative label collection

the annotator was comparing the subject to the target instead of vice versa. Eyewitness descriptions can be influenced by providing a default answer to a question, this is known as anchoring [19]. To avoid anchoring, all drop down boxes were initially void - forcing a response from the annotator.

The users were asked to compare two subjects whilst both were visible. Five subjects were compared to a single target - this simulates the idea of comparing a selection of subjects against a suspect. A single human comparison consists of 16 trait comparisons (shown in table 1), each using one of five comparative labels. The traits chosen were primarily based on MacLeod's work [10]. It can be observed that three traits (gender, ethnicity and skin color) were annotated using absolute labels. These three traits are unsuited to comparative annotations, either due to the inherently categorical nature of the trait or the lack of a suitable comparison criterion. These absolute annotations are not considered when analyzing the comparative annotations and are used only for recognition and retrieval.

The 100 subjects from the Soton gait database were assigned as one of either 20 targets or 80 subjects. Subjects were assigned to users so as to maximize the number of descriptions comparing different subjects and targets.

Performing comparisons between a large group of subjects and a small group of targets also allowed inference of annotations between subjects. If two subjects were both compared against the same target then the comparison between the two subjects could be inferred, reducing the number of comparisons required. Inferring comparisons does, however, introduce errors. If two subjects are both labeled as 'taller' than the target, the inferred comparison would be 'same'. The likelihood is that the subjects are not the same height and we are losing resolution with this assumption. Although lacking precision, this approach allowed us to fully exploit the comparisons we obtained from limited experiments.

In the work reported here, there have been 558 comparisons between subjects and targets collected from 57 annotators.

3.2 Database analysis

Each soft biometric trait comparison is comprised of a single categorical label taken from a set of five ordered labels (see table 1). Each of the five labels are assigned a value, ranging from -2 to 2, based on their order such that -2 represents a 'much less' comparison and +2 a 'much more'. A comparison, C_{st}, between a subject, s, and a target, t, can be described as follows:

$$C_{st} \in \{-2, -1, 0, 1, 2\} \quad (1)$$

The comparative annotations were compared with the absolute categorical labels gathered by Samangooei and Nixon [1]. This comparison between annotation techniques will not show which is better, only how much each technique differs from the other. To determine the difference between the descriptions the comparative label is compared against the absolute labels used to annotate the subject and target. If the absolute labels differ and the comparative label reflects this difference the annotations are recorded as concurring - for example if the target and subject were labeled as 'short' and 'tall' respectively and the comparative descriptor provided was 'taller', we would consider both annotations as concurring. The absolute annotations obviously lack detail; two people labeled as 'tall' are unlikely to be exactly the same height. Thus, small differences can be described using comparative annotations but not absolute labels. In the case of both the subject and target having the same absolute label, the similarity of the comparative annotation cannot be determined. In this case the comparative annotation was recorded as concurring - this ensures we do not overestimate the difference between absolute and comparative annotations. Such that:

$$accuracy(A_t, A_s, C_{st}) = \begin{cases} 1 & A_s < A_t \ and \ C_{st} < 0 \\ 1 & A_s > A_t \ and \ C_{st} > 0 \\ 1 & A_s = A_t \\ 0 & otherwise \end{cases}$$
$$(2)$$

Where A is a value representing an ordered absolute label and C_{st} is a comparison between a target, t, and a subject, s. It was found that the comparative annotations differ from the absolute 17% of the time. This does not necessarily mean that the comparative annotations are better - just that they are different to the absolute labels on 17% of occasions

Figure 5 shows the average difference between absolute and comparative annotations for each trait. The F-ratios, derived by ANOVA analysis, presented within [1] clearly show that absolute labels describe some features better than others. Large differences between absolute and comparative labels for traits

Trait	Description Type	Labels
Arm Length	Comparative	[Much Shorter, Shorter, Same, Longer, Much Longer]
Arm Thickness	Comparative	[Much Thinner, Thinner, Same, Thicker, Much Thicker]
Chest	Comparative	[Much Smaller, Smaller, Same, Bigger, Much Bigger]
Figure	Comparative	[Much Smaller, Smaller, Same, Larger, Much Larger]
Height	Comparative	[Much Shorter, Shorter, Same, Taller, Much Taller]
Hips	Comparative	[Much Narrower, Narrower, Same, Broader, Much Broader]
Leg Length	Comparative	[Much Shorter, Shorter, Same, Longer, Much Longer]
Leg Thickness	Comparative	[Much Thinner, Thinner, Same, Thicker, Much Thicker]
Muscle Build	Comparative	[Much Leaner, Leaner, Same, More Muscular, Much More Muscular]
Shoulder Shape	Comparative	[More Square, Same, More Rounded]
Weight	Comparative	[Much Thinner, Thinner, Same, Fatter, Much Fatter]
Age	Comparative	[Much Younger, Younger, Same, Older, Much Older]
Ethnicity	Absolute	[European, Middle Eastern, Far Eastern, Black, Mixed, Other]
Gender	Absolute	[Female, Male]
Skin Color	Absolute	[White, Tanned, Oriental, Black]
Hair Color	Comparative	[Much Lighter, Lighter, Same, Darker, Much Darker]
Hair Length	Comparative	[Much Shorter, Shorter, Same, Longer, Much Longer]
Neck Length	Comparative	[Much Shorter, Shorter, Same, Longer, Much Longer]
Neck Thickness	Comparative	[Much Thinner, Thinner, Same, Thicker, Much Thicker]

TABLE 1
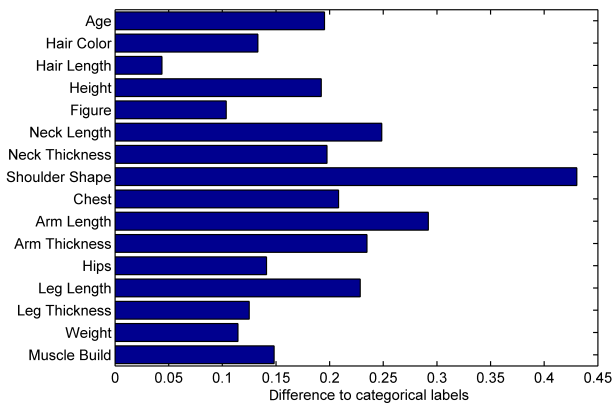Soft traits used to compare subjects

Fig. 5. The average difference of each comparative trait and categorical annotation

demonstrated to be difficult to describe using absolute labels would be indicative of potential improvements when using comparative labels. It can be seen that comparative annotations of arm length (one of the hardest traits to explain categorically) differs on average by 30% compared to absolute labels. Given the inaccuracy of absolute labels for this trait, the difference suggests that the comparative annotations contain more detailed information. Conversely, small differences for traits which were accurately described using absolute annotations, for example hair length, demonstrate that the trait is reliably described using both approaches. It can be observed that the difference between absolute and comparative annotations are on average 5% in respect to hair length, which shows that the comparisons are largely the same as the successful data obtained from the absolute annotations.

4 HUMAN IDENTIFICATION USING COMPARISONS

Comparisons have been introduced as a more robust method for gathering descriptions, but we must consider how they can be applied to identification applications. There are two separate identification situations we will consider in this section. The first recognizes a subject from a database of soft biometric signatures. The second retrieves a subject from a database of videos. In both instances the probe feature vector will be constructed from verbal comparative descriptions.

The first stage in both applications is to convert the comparative descriptions to relative measurements which can be used as a biometric signature. This is described in section 4.1. Sections 4.2 and 4.3 describe identification of subject within a soft biometric database and automatic retrieval of a subject from video footage respectively.

4.1 Relative Measurements

Comparisons are inherently relative; each subject is described using another subject as a benchmark. Comparative annotations must be anchored to convey meaningful subject invariant information. The resulting value is a relative measurement, providing a measurement of a specific trait in relation to the rest of the population. This can be used as a biometric feature allowing retrieval and recognition based on a subject's relative trait measurements.

4.1.1 Law of Comparative Judgment

In 1927, Thurstone introduced the law of comparative judgment [20], allowing the underlying strength of an entity's attribute (also known as the entity's *quality*) to be determined from pairwise comparisons. The model allowed the calculation of *quality scores* for a single pair of entities and was later extended to determine the quality of more than two entities. Thurstone's model employed Gaussian distributions to model pairwise comparisons. It was assumed that an individual's judgment of an entity's quality could be considered as a Gaussian random variable, modeling the subjective nature of assessing 'quality'. Therefore,

the entity's quality score could be modeled by the mean quality of the resulting Gaussian distribution. This section will introduce Thurstone's case V model (which assumes no correlation between compared entities and equal variance of 0.5).

Given two entities, i and j, each considered as a Gaussian random variable. Thurstone states that an individual will compare the entities by drawing two realizations from the entities' quality distributions. The probability that the individual will choose i over j, $P(i > j)$, is dependent on whether their realization of i is greater than their realization of j, such that $P(i > j) = P(i - j > 0)$. Given that $i - j$ is the difference between two Gaussian random variables, $i - j$ is also a Gaussian random variable, where $P(i - j > 0)$ can be calculated using:

$$P(i - j > 0) = \Phi\left(\mu_i - \mu_j\right) \tag{3}$$

where μ is the mean of the corresponding Gaussian random variable and $\Phi(x)$ is the standard normal cumulative distribution function (CDF). Once $P(i > j)$ is determined this can be inverted to find $\mu_i - \mu_j$:

$$\mu_i - \mu_j = \Phi^{-1}(P(i > j)) \tag{4}$$

where Φ^{-1} is the inverse of the standard normal CDF. Equation 4 is known as Thurstone's law of comparative judgment (case V). Obviously, in practical applications, $\mu_i - \mu_j$ is not known and cannot be used to calculate $P(i > j)$, instead $P(i > j)$ must be approximated. Thurstone proposed that the proportion of people who favored entity i over entity j would be an accurate approximation of $P(i > j)$, such that:

$$P(i > j) = \frac{M_{ij}}{M_{ij} + M_{ji}} \tag{5}$$

where M_{ij} is the number of people who favored entity i over entity j.

4.1.2 Elo rating system

To produce relative measurements the comparisons between subjects must be analyzed to identify an ordering within the population in respect to an individual trait. This was achieved using the Elo rating system [21]. In essence, the Elo rating system provides a method of inferring a relative measurement from comparisons and is based on Thurstone's case V model [20]. Elo ratings were designed to quantify the skill of chess players. The performance of a chess player cannot be measured absolutely. Instead the player's (relative) skill level is inferred from matches against other players. This rating system solves a problem very similar to comparative annotations. In soft biometrics the absolute measurements of the traits cannot be directly observed due to the inaccuracy of absolute human descriptions. Instead we can compare traits to infer relative measurements, similar to how chess games compare two players' skill.

In the Elo rating system a 'match' is defined as a comparison between two players, A and B. This comparison could be a chess game or, in the case of soft biometrics, a visual comparison. The outcome of the match is a sample of how the two players differ from each other. The outcome is used to adjust the players' ratings to reflect the sample obtained from the match.

$$R'_A = R_A + K(S_A - E_A) \tag{6}$$

$$R'_B = R_B + K(S_B - E_B) \tag{7}$$

The system adjusts the players' ratings, R, based on the result of match S. The updated rating is derived from the difference between the result of a match, S (1 for a win, 0.5 for a draw and 0 for a loss), and the expected outcome, E, given the players' current ratings. This difference is controlled by K, which defines the maximum rating adjustment resulting from the match.

$$Q_A = 10^{R_A/U} \tag{8}$$

$$Q_B = 10^{R_B/U} \tag{9}$$

$$E_A = \frac{Q_A}{Q_A + Q_B} \tag{10}$$

$$E_B = \frac{Q_B}{Q_A + Q_B} \tag{11}$$

The expected outcome, E, is an adaption of equation 5 based on the Bradley-Terry-Luce model [22], [23], where Q represents a player's current rating. The constant U is chosen to reflect how a player's current rating can affect the expected result.

In chess the unknown measurement is the skill of the chess player - in the case of comparative annotations the unknown variable is the relative measurement of the attribute being compared. Comparisons between subjects provide a measure of difference between the subjects' attributes, just as chess games compare the skill level of the players. This information is used to adjust the inferred relative measurements of the two subjects.

To utilize the Elo rating system for human comparisons a new scoring system (similar to the win-draw-loss system used in chess) is required to compare the expected result to the actual result. Soft biometric traits are compared using five ordered labels, these are assigned a number ranging from -2 to 2 based on their order. The 'score' resulting from a comparison is obtained by normalizing the given label's value to within 0 and 1. If the actual result reflects the expected result the relative measurements are not adjusted. If the actual result disagrees with the expected result, the subjects' relative measurements are adjusted in the direction indicated by the comparison. The size of this adjustment is dependent on the error between the actual and expected results.

In chess the maximum rating adjustment variable, K, can be kept small and over many games the skill

rating of a chess player can be slowly refined. In contrast, our application would benefit from obtaining accurate ratings from the least number of comparisons. This variable can be used to ensure that relative measurements obtained from large numbers of comparisons are comparable to those inferred from just a few comparisons. To perform any form of retrieval or identification the gallery and probe biometric features (i.e. the relative measurements) must be comparable and similar. If K was a constant then the total rating adjustment possible for N comparisons would be $N * K$, this would mean that relative measurements inferred from a small number of comparisons would not be in the same range as those inferred from a large number of comparisons. To solve this K is adjusted based on the number of comparisons available. The maximum rating, M, is used to define $K = M/N$ allowing M to be fully explored by any number of comparisons.

The Elo rating system is used to calculate a single continuous variable, representing the relative strength of an attribute, from visual comparisons. In practice to generate a biometric feature vector describing a suspect, we must first obtain multiple human comparisons - comparing the suspect to multiple subjects (each with predefined Elo ratings). The rating system begins by setting the suspect's Elo ratings (one rating for each comparative trait) to a default value. Each comparison obtained is processed in turn, each time adjusting the suspect's Elo ratings. Once all the comparisons have been considered, a feature vector containing the Elo ratings is constructed. The continuous values within this feature vector, called relative measurements, represent the relative strength of the suspect's physical traits and are used as a biometric signature.

The main advantage of this system is that it does not require exhaustive comparisons between all the subjects to calculate an accurate relative measurement. Instead it adjusts the target's relative measurements based on any available comparisons, taking into account the relative measurements of the compared subjects. In this way the ratings for a set of players can be inferred from a limited set of matches between them.

4.1.3 Accuracy of relative measurements

Relative measurements detail how the subject's traits compare to other subjects within the population. We would expect that the relative measurements, if accurate, would be strongly correlated with the actual physical measurements of the traits. Figure 6 shows the relationship between the relative and actual height measurements. For example, the correlation between pixel height and relative height was statistically significant ($p < 0.0001$) at 0.87 - showing that the relative measurements inferred from human comparisons strongly represent the physical traits. This implies
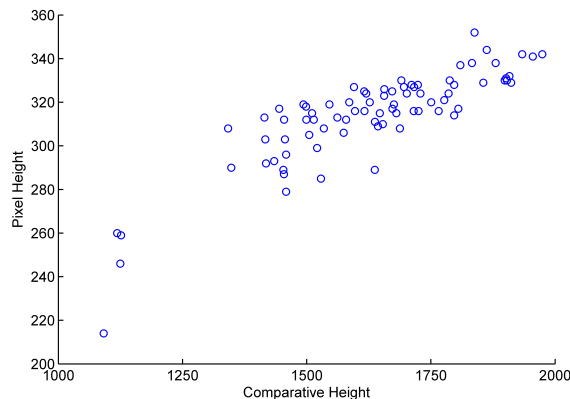

Fig. 6. The relationship between pixel height and relative height

that the Elo rating system has inferred, from visual comparisons, an accurate ordering of the subjects based on height.

The correlation between pixel height and the absolute height labels used previously (figure 2) was found to be 0.71. This is significantly ($p = 0.0018$) weaker than relative measurements mainly due to the highly subjective and categorical nature of the labels.

The relative measurements shown in figure 6 were inferred from all the comparisons in the human comparison database. In application settings we would seek to compare against a minimum number of subjects to achieve an accurate relative measurement. Figure 7 shows the correlation between relative height and pixel height for varying numbers of comparisons per subject. It can be seen that the correlation increases throughout the range presented (1-52 comparisons), clearly demonstrating that additional comparisons improve the accuracy of the resulting relative measurement. The correlation was within 10% of its terminal value after 9 comparisons. We appreciate that this convergence rate perhaps depends on the size of the database used and convergence could require a greater number of comparisons with an increased gallery size. Interestingly the Police and Criminal Evidence Act [24] states that an ideal identity parade should consist of 8 to 12 people, implying that a requirement of 9 comparisons would be suitable for application environments.

4.2 Verbal Identification from Soft Biometric Database

Biometric recognition aims to identify an unknown subject by comparing their biometric signature to a database of biometric signatures. This type of identification is only possible when a database of biometric data is already available. A biometric database could be constructed using previous human comparisons or obtained from other forms of human representation.
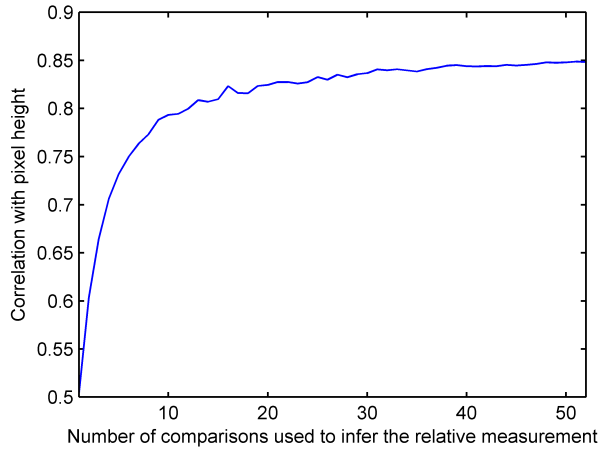
Fig. 7. Correlation between pixel height and relative height with varying amounts of comparisons per subject.

Fig. 8. Verbal identification from soft biometric database

This section will focus on identifying a subject from a soft biometric database formed from previous comparisons whilst section 4.3 will focus on automatic retrieval from video footage.

Soft biometric identification would be ideally suited to criminal investigations where an eyewitness description is available as well as a database of possible suspects each with soft biometric information, in this case obtained from previous human comparisons. The eyewitness would compare the suspect they observed to multiple subjects from the criminal database. Based on the given comparisons a soft biometric feature vector representing the suspect would be inferred and used to query the database. The subjects within the database would be ordered based on their similarity to the feature vector. Figure 8 shows a diagram detailing the identification process. Querying criminal databases using physical descriptions is already common practice within police investigations although currently it is performed using absolute labels and estimates of continuous traits rather than comparative descriptions [25].

The identification experiment aims to retrieve a suspect from an 80 subject database (introduced in section 3.1). The biometric signatures within the database consist of all the 19 traits (table 1), where comparative traits are represented as relative measurements and absolute traits represented as a vector indicating the assigned label. The process starts by selecting a *target* from the database. n randomly sampled comparisons between the target and other subjects will be removed from the database and used to infer the suspect's biometric signature used to query the database (known as the probe). This replicates the eyewitness comparing the suspect to n subjects from the database. n will be varied to investigate how many comparisons are required to retrieve a suspect accurately. The sus-
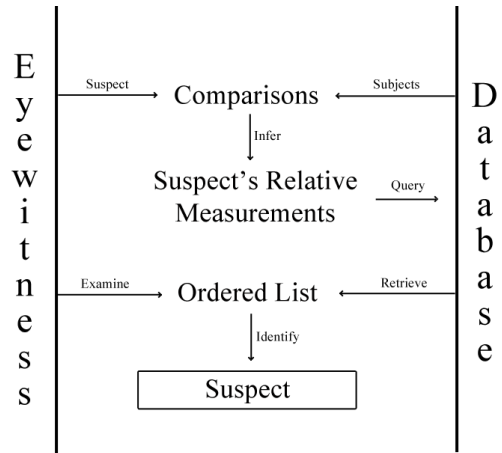
pect's remaining comparisons will be used to produce the biometric signature stored within the database (known as the gallery). The similarity between the probe and every subjects' biometric signature within the database will be assessed using the Euclidean distance metric. The subjects will be ordered based on their similarity to the probe. The position of the suspect's gallery biometric signature within the ordered list shows the retrieval performance of the system. The suspect is correctly identified if their gallery signature is positioned first in the ordered list (known as rank 1). This process will be repeated until the retrieval performance for a certain n remains constant.

The retrieval results shown in this paper are obtained from exhaustively calculating the similarity between the probe and each gallery signature. For larger databases this process could be accelerated by filtering the subjects based on soft biometric features which are reliably and accurately described. It should be noted that in this experiment each annotator has, on average, compared 10 of the 80 subjects within the database. In application settings a single annotator would not describe such a large proportion of the subjects. Given the reduced subjectiveness of comparative descriptions, the comparisons are likely to be the same if more annotators were available and hence we do not believe this biases the experiment.

The recognition accuracy (i.e. rank 1 retrieval accuracy) over varying numbers of probe comparisons (n) is shown in figure 9. The recognition accuracy using just one comparison to construct the probe is 47%. Obviously one comparison only tells us how subjects differ and the resulting relative measurements are very inaccurate. Interestingly this result matches the recognition accuracy when using categorical labels, as seen in figure 3. As more comparisons are received, the accuracy of the probe's relative measurements increase, leading to improved recognition results. It can be seen with 9 comparisons a 91% correct recog-
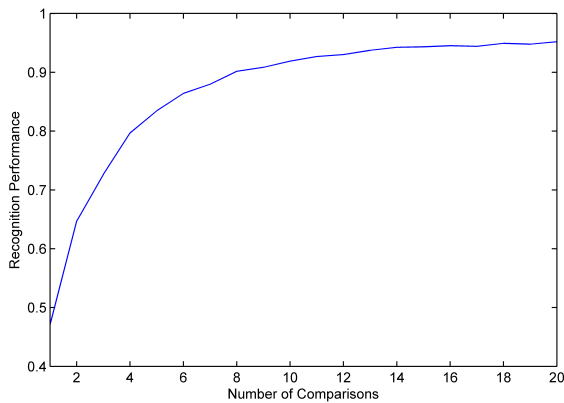
Fig. 9. Recognition accuracy using relative measurements obtained from different numbers of comparisons
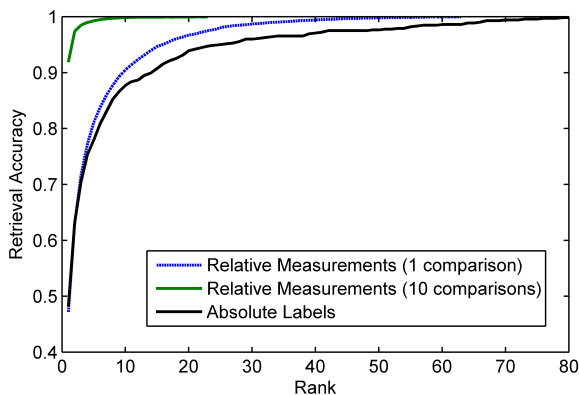

Fig. 10. Retrieval accuracy of absolute labels and relative measurements inferred from 1 comparison and 10 comparisons
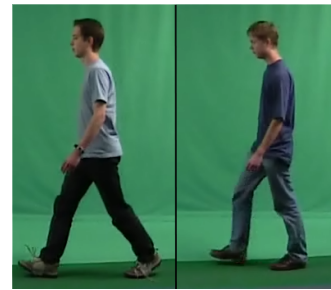

Fig. 11. Incorrect retrieval with 10 comparisons. Left: Database probe. Right: Retrieved subject
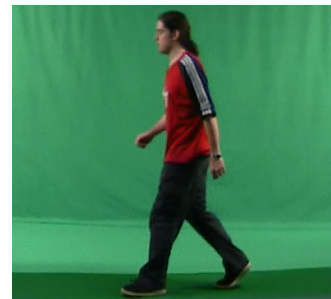
Fig. 12. Subject achieved correct retrieval with only one comparison

nition rate is achieved. This demonstrates that accurate relative measurements are very distinct. The recognition performance continues increasing over the range shown, achieving a 95% correct recognition rate with 20 comparisons. Figure 10 shows the retrieval performance of both relative measurements and absolute labels (absolute labels obtained from [1], more details in section 2.2). Relative measurements inferred from just one comparison outperform absolute labels, achieving a 90% retrieval accuracy at rank 10 (i.e. 90% chance of the suspect being in the first 10 subjects returned from the database) compared to rank 15. As more comparisons are obtained relative measurements vastly outperform absolute labels, achieving a 99% retrieval accuracy at rank 5 with 10 comparisons.

Figure 11 shows an unsuccessful identification. It can be observed that the subjects appear very similar - both having a very similar build, hair length and skin color. The relative measurements of the subjects' traits reflect these similarities resulting in confusion between the two. In comparison, figure 12 shows a subject who was retrieved successfully even with

only one comparison. The male subject has long hair, which is uncommon in the Soton gait dataset, and is also particularly tall. These traits result in a distinct set of relative measurements making retrieval very successful.

As with any biometric system, the scalability of this technique is very difficult to ascertain. In this experiment we have shown that identification is extremely accurate on a database of 80 subjects, we have also shown that increasing the number of comparisons provided also increases the identification performance. It would be reasonable to assume that with larger databases the number of comparisons collected would dictate the accuracy of the system. This provides a mechanism by which the accuracy could be scaled based on the application requirements. It is important to consider that many of the potential applications of such a technique are currently either searched manually (surveillance footage) or utilize crude absolute labels to filter huge databases (criminal databases). Any retrieval system which was able to reduce the number of individuals which required manual evaluation would save money and speed up the process of searching for individuals. It is fair to assume that the advantages of comparative descriptions over absolute labels are a constant and will not vary with database size. For this reason the utilization of comparative descriptions will provide improvements to the filtering currently exploited within the considered applications.

4.3 Verbal Identification from Video Footage

Traditional biometrics identify people by matching biometric signatures. This restricts identification to situations where the subject's biometric signature can be obtained and only permits identification of those subjects whose biometric signature has previously been recorded. Soft biometrics are similar, in that it identifies people by matching signatures. The major difference is that a biometric signature based on relative measurements can be obtained from multiple sources. We have shown how relative measurements can be inferred from human descriptions (section 4.1). Many situations may require the described subject to be identified based on images, surveillance footage, bodily measurements or different biometric signatures. An exciting application of such an approach could be to search surveillance footage around a crime scene for people matching a human description obtained from an eyewitness. This section will introduce how we can deduce relative measurements from visual and biometric representations, focusing on gait signatures. The techniques introduced in this section serve to demonstrate that searching video using human descriptions is possible - justifying future research within this area.

To reliably determine relative measurements from a visual representation of a human, the representation must:

- be a standard signature allowing comparisons between different individuals.
- contain visual features which relate to or describe the soft features we are trying to determine.

Biometric signatures are a suitable form of representation as they focus on producing visual features which are consistent and comparable between different individuals. To accurately determine relative measurements the human representation must contain information about the soft traits (shown in table 1) which compose the soft biometric signature. In this case we are using mostly bodily features, making gait signatures the ideal biometric to use for surveillance applications. For these reasons gait signatures were studied within this research. It is important to note that any visual or biometric signature which encompassed the traits being described could be used.

Average silhouette gait signatures describe the averaged summation of a subject's binary silhouettes across one gait cycle [26]. This representation encompasses the physical features which relate to the traits presented within table 1, which was believed to allow accurate regression. The unwrapped silhouette gait signature proposed by Wang et al. was also implemented [27]. The advantage of this signature is that many of the physical measurements described within the soft traits are explicitly measured (in terms of pixel distance from centroid) within the gait signature rather than being implicit within the pixel data. A
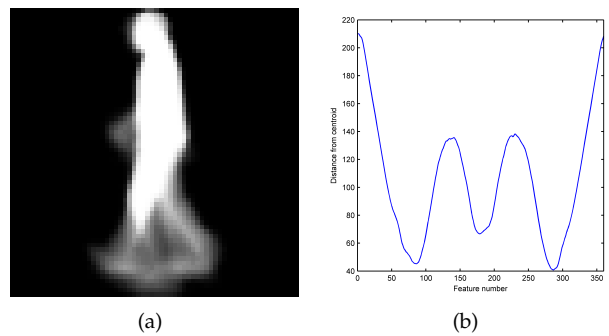

(a)   (b)

Fig. 13. Example gait signatures for the subject shown in figure 12. a) Average gait signature b) Unwrapped silhouette gait signature

comparison of the two gait signatures can be seen in figure 13. A third gait signature, fusing both the average and unwrapped gait signatures into one feature vector, was also investigated.

The current scenario shows the consistency between measures derived from human labeling and those derived by automated analysis of gait data. The gait data is laboratory based side view data as used in many early approaches to gait biometrics, we demonstrate that it is possible to derive this set of relative measures from this gait data, and this set is much extended from other approaches such as those which focus on gender alone. Application data requires translation to other viewpoints and that represents an extension of the material presented here in exactly the same way that viewpoint invariant gait biometrics was an extension to studies previously conducted in the planar view [28].

Automated gait analysis could be performed by directly measuring features present within the gait signature. By measuring features we are assuming we understand the precise meaning behind the label. Automated methods can be used to detect the correlations between visual features and their associated labels, especially if a model-based approach is incorporated in the procedure.

A support vector machine (SVM) was used to predict a relative measurement given a gait signature. The gait signatures were represented as a vector of values representing either pixel intensities (average gait signature) or pixel distance from centroid (unwrapped silhouette gait signature). The relative measurements were normalized to a range of 0-1 and used to train the SVM. Each trait was considered in turn and the system was validated using 10-fold cross validation. The three absolute categorical labels within the soft biometric signature (see table 1) were predicted using a SVM configured for classification rather than regression.

Figure 14 shows the correct classification rate of the categorical labels generated automatically from gait signatures. Both average and unwrapped gait

Gait Signature	Comparative features	Absolute features
Average	13.9%	26.7%
Unwrapped	13.4%	26.7%
Average+Unwrapped	12.8%	16.2%

TABLE 2
Errors present in features obtained automatically from
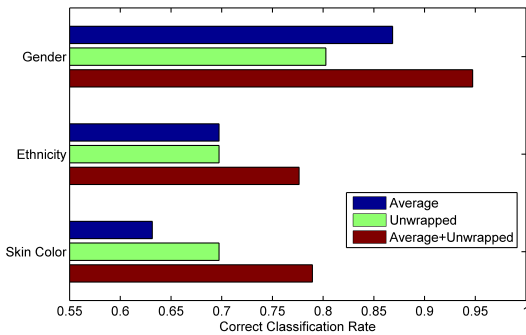different gait signatures

Fig. 14. Correct classification rate of absolute labels

signatures achieved a 73.3% average correct classification rate (CCR). It can be seen that gender is very successfully classified. Gender has been shown to be highly correlated with other physical features [29] (especially height) which makes it easy to determine from both gait signatures. Skin color and ethnicity both rely on information which is not encoded within the signatures resulting in poorer classification performance. The Soton gait database is largely composed of Caucasian individuals, making up 70% of the database. Given the little visual information available to classify these traits, it is optimal to classify all subjects as the mode class, hence the roughly 70% accuracy. The fusion signature exceeds the CCR of the other two signatures resulting in a CCR of 83.8%.

Predicting gender from gait signatures is a growing research area with interest in bolstering gait biomet-
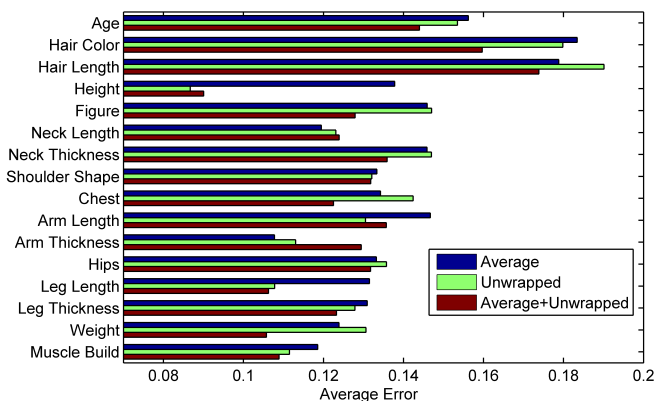

Fig. 15. Errors present in relative measurements obtained automatically from different gait signatures

rics with additional soft traits. Obviously the aim of our research differs from the research within this area, although it does offer a performance comparison for predicting this particular trait. Hu et. al. [30] utilize Gabor wavelets to obtain low level features of gait silhouettes. Gender classification is achieved using Hidden Markov Models. A 96.77% correct classification rate was obtained when predicting gender from silhouettes recorded from a side on viewpoint. Yoo et. al. [31] utilized SVMs trained to classify gender based on model based gait signatures obtained from the Soton gait database. A 96% correct classification rate was achieved. In this paper we have achieved a 94.7% correct classification rate when predicting gender from a gait signature composed of both average and unwrapped signatures.

Figure 15 shows the errors present within relative measurements generated automatically from different gait signatures. The relative measurements used to train the SVM were all normalized to the range 0-1 and as such the accuracy was the amount of error between the training and the generated relative measurements expressed in proportion. The average error rates of the different gait signatures can be seen in table 2. It can be observed that the unwrapped signature performed slightly better than the average gait signature. Improvements of 4% (a 0.6% error reduction) were found by fusing both the average and unwrapped gait signatures into one feature vector, exploiting the advantages of both signatures.

To determine the application potential of such a system, we must also consider the retrieval accuracy. The retrieval process is identical to that introduced in section 4.2 although all the subjects' soft biometric feature vectors within the gallery are generated automatically from the fused average and unwrapped gait signatures.

Based on figures 14 and 15 it can be seen that some traits are more reliably predicted from gait signatures. When considering retrieval, these accurate traits should be favored. Feature set selection was used to weight the importance of each trait in the retrieval process. The feature weights were defined using a genetic algorithm based feature set selection, and the resulting weights used to bias the euclidean distance which defines the similarity between two biometric signatures (probe and gallery in this case).

The resulting retrieval performance can be seen in figure 16. Although the retrieval accuracy is only 20% at rank 1, it quickly increases achieving a 69% retrieval accuracy at rank 9 and a 90% at rank 19. This result highlights the possibility of *automatically* filtering video data based on a description. Improvements may be found with the use of model-based gait signatures, which would provide a stronger relationship between the gait signatures and soft biometric relative measurements. This would increase the accuracy of the automatically generated relative measurements
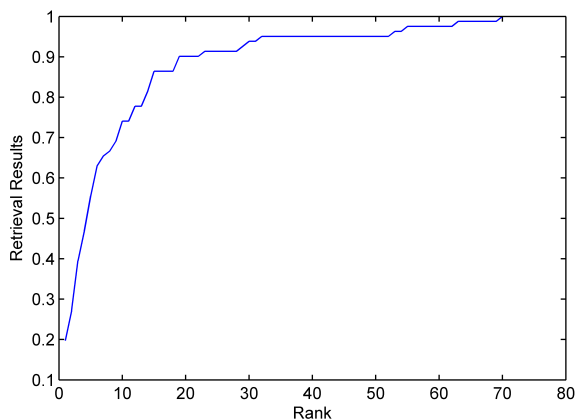
Fig. 16. Retrieval accuracy using relative measurements obtained automatically from unwrapped+average gait signatures

leading to improved retrieval results. We believe these results represent a good start to this difficult problem. Future research will aim to improve the retrieval performance.

5 DISCUSSION AND CONCLUSIONS

Soft biometrics allow identification based on natural human descriptions of behavioral and physical traits. Identification is only possible if the description is accurate and detailed. Conventional forms of human description focus on absolute labels and estimations which can be unreliable due to their highly subjective nature and the errors common with continuous estimations. Visual comparisons composed of categorical labels have been introduced as an alternative - increasing the objectivenesss of labeled descriptions whilst inferring informative and robust continuous measurements of the subject's traits.

Comparisons between subjects from the Soton gait database were collected. Each annotator was asked to compare a single target to multiple subjects. It was found that comparative descriptions differed on 17% of occasions when compared against absolute categorical descriptions. Comparative annotations of traits which were poorly described using absolute labels were found to differ by up to 40%, suggesting the comparative annotations contained new and more detailed information.

Human comparisons must be anchored before they can be used as a biometric signature. The Elo rating system was used to infer an ordering between subjects in respect to a specific trait. This produced a relative measurement of the trait in respect to the rest of the population. The relative measurements were shown to strongly represent actual trait measurements; results comparing actual height to the inferred relative height showed a correlation of 0.87.

Biometric signatures consisting of 19 relative measurements were exploited to allow biometric identification. Recognition of subjects from a soft biometric database demonstrated the discriminative power of relative measurements, achieving a recognition accuracy of 92% with ten comparisons. This increased to 95% with 20 comparisons.

Relative measurements were also obtained automatically from video footage using a support vector machine, allowing video data to be searched automatically using a human description. A retrieval accuracy of 90% was achieved at rank 19, permitting a database of videos to be filtered automatically. Future research will attempt to improve the retrieval results allowing automatic identification from video footage.

Comparative descriptions have been shown to contain more discriminative information and present an innovative approach to obtaining robust human descriptions for soft biometrics and possibly eyewitness descriptions.

REFERENCES

[1] S. Samangooei and M. S. Nixon, "Performing Content-based Retrieval of Humans using Gait Biometrics," *Multimedia Tools and Applications*, vol. 49, no. 1, pp. 195–212, 2010.
[2] H. Ailisto, M. Lindholm, S. M. Makela, and E. Vildjiounaite, "Unobtrusive user identification with light biometrics," in *Proc. NordiCHI*, 2004, pp. 327–330.
[3] A. K. Jain, K. Nandakumar, X. Lu, and U. Park, "Integrating faces, fingerprints, and soft biometric traits for user recognition," in *BioAW*, vol. LNCS 3087, 2004, pp. 259–269.
[4] U. Park and A. K. Jain, "Face Matching and Retrieval Using Soft Biometrics," *IEEE Trans on Information Forensics and Security*, vol. 5, no. 3, pp. 406–415, Sep. 2010.
[5] A. Dantcheva, J. Dugelay, and P. Elia, "Soft biometrics systems: Reliability and asymptotic bounds," in *BTAS*, 2010, pp. 1–6.
[6] E. F. Loftus, *Eyewitness testimony*. Harvard U. Pr., 1996.
[7] C. A. Meissner, S. L. Sporer, and J. W. Schooler, "Person descriptions as eyewitness evidence," *Handbook of eyewitness psychology*, vol. 2, pp. 3–34, 2007.
[8] J. C. Yuille and J. L. Cutshall, "A case study of eyewitness memory of a crime," *Journal of Applied Psychology*, vol. 71, no. 2, pp. 291–301, 1986.
[9] L. L. Kuehn, "Looking down a gun barrel: Person perception and violent crime." *Perceptual and Motor Skills*, vol. 39, no. 3, pp. 1159–1164, 1974.
[10] M. D. MacLeod, J. N. Frowley, and J. W. Shepherd, "Whole body information: Its relevance to eyewitnesses," in *Adult eyewitness testimony: Current trends and developments*. Cambridge University Press, 1994, ch. 6.
[11] R. H. Flin and J. W. Shepherd, "Tall Stories: Eyewitnesses' Ability to Estimate Height and Weight Characteristics," *Human Learning: Journal of Practical Research & Applications*, vol. 5, no. 1, pp. 29–38, 1986.
[12] D. A. Reid, S. Samangooei, C. Chen, M. S. Nixon, and A. Ross, "Soft Biometrics for Surveillance: An Overview," in *Handbook of statistics*. Elsevier, 2013, vol. 31, ch. 13, pp. 327–351.
[13] J. Shutler, M. Grant, M. S. Nixon, and J. N. Carter, "On a large sequence-based human gait database," in *Proc RASC*. Springer Verlag, 2002, pp. 66–72.
[14] S. Denman, C. Fookes, A. Bialkowski, and S. Sridharan, "Soft-Biometrics: Unconstrained Authentication in a Surveillance Environment," *Digital Image Computing: Techniques and Applications*, pp. 196–203, 2009.
[15] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and Simile Classifiers for Face Verification," in *ICCV*, 2009, pp. 365–372.

[16] G. Wang, D. Forsyth, and D. Hoiem, "Comparative object similarity for improved recognition with few or no examples," in *CVPR*. IEEE, 2010, pp. 3525–3532.

[17] D. Parik and K. Grauman, "Relative Attributes," in *ICCV*, 2011.

[18] T. Joachims, "Optimizing search engines using clickthrough data," in *SIGKDD*. ACM, 2002, pp. 133–142.

[19] G. B. Chapman and E. J. Johnson, "Incorporating the irrelevant: Anchors in judgments of belief and value." *Heuristics and Biases: The Psychology of Intuitive Judgment*, pp. 120–138, 2002.

[20] L. L. Thurstone, "A law of comparative judgment." *Psychological Review; Psychological Review*, vol. 34, no. 4, p. 273, 1927.

[21] A. E. Elo, *The rating of chessplayers, past and present*. Batsford, 1978.

[22] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. The method of paired comparisons," *Biometrika*, vol. 39, no. 3/4, pp. 324–345, 1952.

[23] R. D. Luce, *Individual choice behavior*. John Wiley, 1959.

[24] M. Zander, *The Police and Criminal Evidence Act 1984*. Sweet & Maxwell, 2005.

[25] National Policing Improvement Agency, *PNC User Manual, Volume 2*, Nov. 2009.

[26] Z. Liu and S. Sarkar, "Simplest representation yet for gait recognition: Averaged silhouette," in *ICPR*, vol. 4. IEEE, 2004, pp. 211–214.

[27] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *TPAMI*, vol. 25, no. 12, pp. 1505–1518, 2003.

[28] M. Goffredo, I. Bouchrika, J. N. Carter, and M. Nixon, "Self-Calibrating View-invariant Gait Biometrics," *Transactions on Systems, Man, and Cybernetics–Part B: Cybernetics*, 2010.

[29] D. A. Reid and M. S. Nixon, "Imputing Human Descriptions in Semantic Biometrics," in *Proceedings of the 2nd ACM workshop on Multimedia in forensics, security and intelligence*, 2010, pp. 25–30.

[30] M. Hu, Y. Wang, and Z. Zhang, "Maximisation of mutual information for gait-based soft biometric classification using Gabor features," *IET biometrics*, vol. 1, no. 1, pp. 55–62, 2012.

[31] J. Yoo, D. Hwang, and M. S. Nixon, "Gender classification in human gait using support vector machine," *Lecture notes in computer science*, vol. 3708, p. 138, 2005.

Mark Nixon is the Professor in Computer Vision at the University of Southampton UK. His research interests are in image processing and computer vision. His team develops new techniques for static and moving shape extraction which have found application in automatic face and automatic gait recognition and in medical image analysis. His team were early workers in face recognition, later came to pioneer gait recognition and more recently joined the pioneers of ear biometrics. Amongst research contracts, he was Principal Investigator with John Carter on the DARPA supported project Automatic Gait Recognition for Human ID at a Distance and he was previously with the FP7 Scovis project and is currently with the EU-funded Tabula Rasa project. His vision textbook, with Alberto Aguado, Feature Extraction and Image Processing (Academic Press) reached 3rd Edition in 2012 and has become a standard text in computer vision. With Tieniu Tan and Rama Chellappa, their book Human ID based on Gait is part of the Springer Series on Biometrics and was published in 2005. He has chaired/ program chaired many conferences (BMVC 98, AVBPA 03, IEEE Face and Gesture FG06, ICPR 04, ICB 09, IEEE BTAS 2010) and given many invited talks. Mark is a member of IAPR TC4 Biometrics and of the IEEE Biometrics Council. Dr. Nixon is a Fellow IET and FIAPR.

Sarah Stevenage is a cognitive psychologist working in the field of identity and identification. Following a degree and PhD in cognitive psychology at the University of Exeter, UK, she now works as a Senior Lecturer and Associate Dean at the University of Southampton. Her interests extend into the forensic psychology domain, with particular interests in novel forms of biometric identification, together with the linkage between real world and digital identity footprints.

Daniel Reid is a PhD student at the University of Southampton researching into human identification using soft biometrics. His research interests include computer vision, statistics and machine learning. Prior to his PhD, Daniel studied at the University of Reading achieving a Masters in Cybernetics and Artificial Intelligence.