

# The Use of Semantic Human Description as a Soft Biometric

Sina Samangooei

Baofeng Guo

Mark S. Nixon

**Abstract**—Gait as a biometric has a unique advantage that it can be used when images are acquired at a distance and other biometrics are at too low a resolution to be perceived. In such a situation, there is still information which can be readily perceived by human vision, yet is difficult to extract automatically. We examine how this information can be used to enrich the recognition process. We call these descriptions *semantic annotations* and investigate their use in biometric scenarios. We outline a group of visually assessable *physical traits* formulated as a mutually exclusive set of *semantic terms*; we contend that these traits are usable in soft biometric fusion. An experiment to gather semantic annotations was performed and the most reliable traits are identified using ANOVA. We rate the ability to correctly identify subjects using these semantically prescribed traits, both in isolation as well as in fusion with an automatically derived gait signature.

---

*“HOLOFERNES: Novi hominem tanquam te: his humour is lofty, his discourse peremptory, his tongue filed, his eye ambitious, his gait majestic, and his general behavior vain, ridiculous, and thrasonical. He is too picked, too spruce, too affected, too odd, as it were, too peregrinate, as I may call it.”*

– William Shakespeare, *Love’s Labour’s Lost*

---

## I. INTRODUCTION

In today’s security conscious global climate there is an increasing need for effective human identification. When close contact and subject co-operation is assured, biometric techniques using DNA, iris signature and fingerprint recognition have been shown to address this need. Non-contact biometrics such as gait [22], face [26] and ear [10] address the need for identification and verification when subjects are non co-operative and without contact. We consider how these processes can be enriched by human descriptions.

The human ability to identify individuals has been shown to be consistently effective at a distance [30], under varying weather conditions, light conditions [28] and behavioural configurations [3]. Notwithstanding Shakespeare’s lofty observations; humans can efficiently use higher level *semantic* concepts such as Sex [16], Race [1] and Build [18] to aid description and identification. However, human recognition has various issues (e.g. memory limitations and descriptive completeness) which can impede accurate description ability,

recall and subsequently recognition. It is clear that, for the purposes of improved identification, the potential affordances offered by automated techniques compared to human descriptions are distinct and indeed complementary.

The primary goal of this paper is to outline the important labels used by humans for description of physical characteristics of other humans at a distance. These qualitative annotations or *semantic terms* are defined as words or phrases portraying directly interpretable meaning to annotators when describing some visually discernible *physical trait* of a subject. We present a set of experiments which explore the capability of these semantic terms for identification purposes, both in isolation and in fusion with a gait biometric signature. Gait was chosen as the primary biometric for the fusion task as it is one of the only biometrics which can achieve identification at a distance and therefore portrays capabilities comparable to those of human perception. Furthermore, semantic traits noticeable at a distance complement static features of gait and therefore allow further studies beyond fusion, such as the *correlation* between semantic and automatic gait signatures.

The rest of this paper is organised in the following way. Section II explores and justifies the physical traits and semantic terms to be used for annotation. Section III outlines the source of semantic data including the measures undertaken to counteract weaknesses in human description. Automatic gait features used are also briefly summarised. In Section IV we show the comparative recognition rates of semantic features, automatic features and various fusion techniques combining the two. Finally in Section V, we discuss what has been achieved and future work.

## II. HUMAN DESCRIPTIONS AS A BIOMETRIC

One of the first attempts to systematically identify people based on their physical traits was the anthropometric system developed by Bertillon [2] in 1896. His system used eleven precisely measured traits of the human body including height, length of right ear and width of cheeks. Although his approach was recognised in the field of criminology for its use in filtering lists of suspects, it was quickly superseded by other forms of forensic analysis such as fingerprints.

An early mention of the use of anthropomorphic traits for aiding primary biometric schemes was suggested by Wayman [31] in the form of filtering by age or gender. One of the few explorations into this approach was performed later by Nandakumar et al. [21], who used methods for automatic extraction of “soft” biometric values and fusion methods of these features with “primary” biometrics using a Bayesian framework. His experiments showed an improvement of

---

All the authors are with the Information: Signals, Images, Systems research group, a part of the School of Electronics and Computer Science at the University of Southampton, United Kingdom [ss06r,bg,msn]@ecs.soton.ac.uk

around 1-2% when combining ethnicity and gender traits with fingerprint signals. Other related approaches such as Ze-wail et al. [34] use iris colour (a soft biometric) with fingerprint and iris using a weighted average scheme and a Parzen Classifier. These approaches used automatically extracted soft biometrics from existing video or image signals.

No existing work has attempted the fusion of manually annotated traits from several human sources as a soft biometric. The key difference between semantic biometrics and existing soft biometrics is the measurement scheme that is used. Soft biometric techniques use a mixture of categorical metrics (e.g. Race) and value metrics (e.g. Height) to formulate their features. Humans are generally less consistent when making value judgements in comparison to category judgements. Subsequently, in our approach we formulate all traits with sets of mutually exclusive semantic terms rather than using value metrics. This approach is more representative of the categorical nature of human cognition [29] [19] [30]. The following sections outline the traits chosen and their associated semantic terms.

### *Physical Traits*

This section outlines the factors taken into consideration when choosing the exact physical traits to investigate. To match the capabilities of automatic gait biometrics, one of our primary concerns was to choose traits that are discernible by humans at a distance. To do so we must firstly ask what traits individuals *consistently* and *accurately* notice in each other at a distance. Three independent traits - Age, Race and Sex, are agreed to be of primary significance in cognitive psychology [19]. In human recognition of faces, Age, Sex and Race are also amongst the main “visually derived semantic categories” noticed in unfamiliar faces [3] [23]. In gait recognition, humans have been shown to successfully perceive Age and Sex using generated point light experiments [30] [20] with limited visual cues. Other factors such as the target’s perceived somatotype [19] (build or physique attributes) are also prominent in cognition.

The eyewitness testimony research community have a relatively mature idea of which concepts witnesses are more likely to recall when describing individuals [32]. Koppen and Lochun [15] provide an investigation into witness descriptions in archival crime reports. Not surprisingly, the most accurate and highly mentioned traits were Sex (95% mention 100% accuracy), Height (70% mention 52% accuracy), Race (64% mention 60% accuracy) and Skin Colour (56% mention, accuracy not discussed). Detailed head and face traits such as Eye Shape and Nose Shape are not mentioned as often and when they are mentioned, they appear to be inaccurate. More prominent head traits such as Hair Colour and Length are mentioned more consistently, a result also noted by Yarmey and Yarmey [33]. Descriptive features which are visually prominent yet less permanent (e.g. clothing) often display instability over time and are of less interest than other more permanent physical traits.

### *Build Traits*

Traits regarding build are of particular interest, having a clear relationship with gait while still being reliably recalled by eyewitnesses at a distance. Few studies thus far have attempted to explore build in any amount of detail beyond the brief mention of Height and Weight. MacLeod et al. [18] performed a unique analysis on whole body descriptions using bipolar scales to define traits. There were two phases in their approach towards developing a set of descriptive build traits.

Firstly a broad range of useful descriptive traits were outlined with a series of experiments where a mixture of moving and stationary subjects were presented to a group of annotators who were given unlimited time to describe the individuals. A total of 1238 descriptors were extracted, of which 1041 were descriptions of overall physique and the others were descriptions of motion. These descriptors were grouped together (where synonymous) and a set of 23 traits generated, each formulated as a bipolar five-point scale.

In the second phase the reliability and descriptive capability of these traits were gauged. Annotators were asked to watch video footage of subjects walking at a regular pace around a room and rate them using the 23 traits identified. The annotators were then split into two groups randomly from which two mean values were extracted for each subject for each trait. The product moment correlation (Pearson’s  $r$ ) was calculated between the sets of means and was used as an estimate of the reliability for each trait. A principal components analysis was also used to group traits which represented similar underlying concepts. The 13 most reliable terms and most representative of the principle components have been incorporated into our final trait set.

Jain et al. [12] outline a set of key characteristics which determine a physical trait’s suitability for use in biometric applications, these include: Universality, Distinctiveness, Permanence and Collectability

The choice of our physiological traits keeps these tenets in mind. Our semantic descriptions are universal in that we have chosen factors which everyone has. We have selected a set of subjects who appeared to be semantically distinct in order to confirm that these semantic attributes can be used. The descriptions are relatively permanent: overall skin colour naturally changes with tanning, but our description of skin colour have racial overtones and these are perceived to be more constant. Our attributes are easily collectible and have been specifically selected for being easily discernible at a distance by humans. However much care has been taken over procedure and definition to ensure consistency of acquisition.

Using a combination of the studies in cognitive science, witness descriptions and the work by MacLeod et al. [18] we created a list of visual physiological traits, presented in Table I.

### *Semantic Terms*

All the traits to be captured were annotated from a limited set of mutually exclusive terms. This is naturally achieved for certain traits, primarily when no applicable underlying

value order exists (Sex, Hair Colour etc.). For other traits representable with intuitive value metrics (Age, Lengths, Sizes etc.) bipolar scales representing concepts from *Small* to *Large* are used as semantic terms. This approach closely matches human categorical perception. Annotations obtained from such approaches have been shown to correlate with measured numerical values [5].

Perhaps the most interesting trait for which to define a limited set of terms was Race. A large corpus of research [8] [24] [1] explores racial classification, many works outline different Racial categories; ranging from the use of 3 to 200 distinct races, non necessarily convergent. Indeed, experts are still divided over the basic principal of whether Race is in fact real at all<sup>1</sup>. Furthermore, Race perception is especially affected by subject variables, portraying instability between global regions. This is exemplified by the 2000 US census<sup>2</sup> which included the races American Indian, Pacific Islander and Other (including hispanic) where the 2001 UK census<sup>3</sup> contained none of these but did include such races as Indian, Pakistani, Bangaldeshi and White Irish.

It is however undeniable that physical Race categorisations are readily attempted by humans during recognition, prompted by some global combination of: Skin Colour; Hair Colour; Build; Facial Features etc. The overwhelming amount of contradictory Race specific research itself could be taken as evidence for the significance of Race. Evidence also exists in the psychology community [23] [3] and the witness analysis community [15] as well as many practical identification applications such as human description forms [11]. Therefore, our Race terms encompass the three categories mentioned most prominently in the literature but also include an extra two categories (Indian and Middle Eastern) localising our categories to the UK.

### III. DATA ACQUISITION

In this section we describe the source of the data gathered using our semantic and automatic features. Our subjects are 10 individuals each with a minimum of 6 video samples from the Southampton University Gait Database [27], chosen for their distinctiveness in the semantic visual sense. The videos involve the subject walking at a natural pace, side on to the camera view.

#### *Semantic Features*

Semantic annotations were collected using the GaitAnnotate system; a web application designed to show arbitrary biometric data sources to users for annotation, as shown in Fig. 1. This interface allows annotators to view all video samples of a subject as many times as they require. Annotators were asked to describe subjects by selecting semantic terms for each physical trait. They were instructed to label *every* trait for *every* subject and that each trait should be completed with the annotator's own notions of what the trait *meant*. Guidelines were provided to avoid common

<sup>1</sup><http://www.pbs.org/wgbh/nova/first/race.html>

<sup>2</sup><http://www.census.gov/main/www/cen2000.html>

<sup>3</sup><http://www.ons.gov.uk/census/>

TABLE I  
PHYSICAL TRAITS AND ASSOCIATED SEMANTIC TERMS

Body Shape		
1. Arm Length	[Very Short, Short, Average, Long, Very Long]	
2. Arm Thickness	[Very Thin, Thin, Average, Thick, Very Thick]	
3. Chest	[Very Slim, Slim, Average, Large, Very Large]	
4. Figure	[Very Small, Small, Average, Large, Very Large]	
5. Height	[Very Short, Short, Average, Tall, Very Tall]	
6. Hips	[Very Narrow, Narrow, Average, Broad, Very Broad]	
7. Leg Length	[Very Short, Short, Average, Long, Very Long]	
8. Leg Shape	[Very Straight, Straight, Average, Bow, Very Bowed]	
9. Leg Thickness	[Very Thin, Thin, Average, Thick, Very Thick]	
10. Muscle Build	[Very Lean, Lean, Average, Muscly, Very Muscly]	
11. Proportions	[Average, Unusual]	
12. Shoulder Shape	[Very Square, Square, Average, Rounded, Very Rounded]	
13. Weight	[Very Thin, Thin, Average, Fat, Very Fat]	
Global		
14. Age	[Infant, Pre Adolescence, Adolescence, Young Adult, Adult, Middle Aged, Senior]	
15. Race	[Other, European, Middle Eastern, Far Eastern, Black, Mixed]	
16. Sex	[Female, Male]	
17. Skin Colour	[White, Tanned, Oriental, Black]	
Head		
18. Facial Hair Colour	[None, Black, Brown, Blond, Red, Grey]	
19. Facial Hair Length	[None, Stubble, Moustache, Goatee, Full Beard]	
20. Hair Colour	[Black, Brown, Blond, Grey, Red, Dyed]	
21. Hair Length	[None, Shaven, Short, Medium, Long]	
22. Neck Length	[Very Short, Short, Average, Long, Very Long]	
23. Neck Thickness	[Very Thin, Thin, Average, Thick, Very Thick]	

confusions e.g. that height of an individual should be assigned absolutely compared to perceived global "Average" where traits such as "Arm Length" could be annotated in comparison to the subject's overall physique. This annotation data was gathered from subjects present in the video set, as well as from subjects not present (e.g. a class of Psychology students).

To gauge an upper limit for the capabilities of semantic data we strive to assure our data is of optimal quality. The annotation gathering process was designed carefully to avoid (or allow the future study of) inherent weaknesses and inaccuracies present in human generated descriptions. The error factors that the system was designed to deal with include:

- **Memory[6]** - Passage of time may affect a witness' recall of a subject's traits. Memory is affected by variety of factors e.g. the construction and utterance

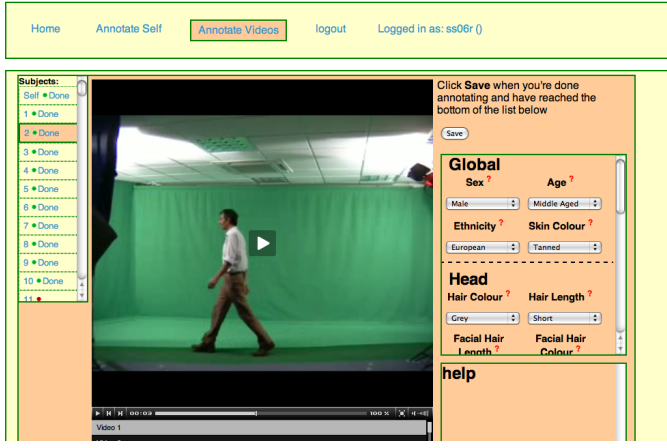


Fig. 1. Example of GAnn interface

of featural descriptions rather than more accurate (but indescribable) holistic descriptions. Such attempts often alter memory to match the featural descriptions.

- **Defaulting[17]** - Features may be left out of descriptions in free recall. This is often not because the witness failed to remember the feature, but rather that the feature has some default value. Race may be omitted if the crime occurs in a racially homogenous area, Sex may be omitted if suspects are traditionally Male.
- **Observer Variables[7][23]** - A person's own physical features, namely their self perception and mental state, may affect recall of physical variables. For example, tall people have a skewed ability to recognise other tall people but will have less ability when it comes to the description shorter individuals, not knowing whether they are average or short.
- **Anchoring[4]** - When a person is asked a question and is initially presented with some default value or even seemingly unrelated information, the replies given are often weighted around those initial values. This is especially likely when people are asked for answers which have some natural ordering (e.g. measures of magnitude)

We have designed our semantic data gathering procedure to account for all these factors. Memory issues are addressed by allowing annotators to view videos of subjects as many times as they please, also allowing them to repeat a particular video if necessary. Defaulting is avoided by explicitly asking individuals for each trait outlined in Table I, this means that even values for apparently *obvious* traits are filled in and captured. This style of interrogative description, where constrained responses are explicitly requested, is more complete than free-form narrative recall but may suffer from inaccuracy, though not to a significant degree [33]. Subject variables can never be completely removed so instead we allow the study of differing physical traits across various annotators. Users are asked to self annotate based on self perception, also certain subjects being annotated are themselves annotators. This allows for some concept of the annotator's

own appearance to be taken into consideration when studying their descriptions of other subjects. Anchoring can occur at various points of the data capture process. We have accounted for anchoring of terms gathered for individual traits by setting the default term of a trait to a neutral "Unsure" rather than any concept of "Average".

To allow for efficient analysis after the annotation data was gathered, a process was undertaken to numerically represent them. A *Semantic Feature Vector* is generated from each annotation of each subject by giving each physical trait a single normalised numeric value proportional to the rank or order of the term it was assigned<sup>4</sup>. For example the trait "Height" is given a low value (e.g.  $1/5 = 0.2$ ) to represent the semantic term "Very Short" and a high value (e.g.  $5/5 = 1$ ) to represent "Very Tall". This representation has the ability to show intermediate values between qualitative terms, for example if someone is somewhere in between "Average" and "Tall" due to the responses of several different annotators.

For each of the 38 annotators a response is returned for each of the 10 subjects for each of the 23 traits. If an annotator responds as "Not Visible" for a trait, or does not provide the annotations at all, their response is set to the numeric mean of that trait across all annotators across all subjects. This results in a complete  $38 \times 10 \times 23$  data structure whose manipulation for analysis is discussed in further detail where appropriate in Section IV.

#### Automatic Gait Features

The primary biometric feature vectors used for these experiments were symmetry gait signatures, as originally developed by Hayfron-Acquah et al. [9]. Symmetry signatures are generated using a discrete symmetry transform which has the ability to estimate symmetry of an image without knowledge of its exact shape. For each gait video, first the subject is extracted from the scene with a median background subtraction, transformed into a binary silhouette and edge detected using the Sobel operator. The edges are then used to extract the symmetry map of each frame. The gait signature is the averaged summation of all these symmetry maps across one gait cycle.

This gait signature may contain noise and so its intensity values are not used directly as a feature vector. Rather, a 2D Fourier decomposition is applied which exposes the frequency components. A low-pass filter is then applied by zeroing all frequency components more than 15 units radially. This filter annuls high frequency components which are likely to represent noise rather than useful subject information, more likely contained in low frequencies. The cut off radius of 15 was suggested empirically by Hayfron-Acquah et al. [9]. The logarithm of the magnitude components within this radius are each taken as values along orthogonal dimensions and represent the feature vector of the particular signature.

<sup>4</sup>The rank of the term is defined as its position due to some inherent or arbitrary order with regard to other possible semantic terms of a particular trait

This results in 709 automatic feature components which describe each sample of each of the 10 subjects. Manipulation of this data is discussed in further detail in Section IV.

#### IV. EXPERIMENTS

In this section we outline a set of experiments used to explore the semantic annotation data we have designed and collected. We present results showing the most significant semantic traits and performance of the semantic features in a recognition task, both isolation and in feature and score fusion.

##### A. Semantic Feature Subset Selection

A one-way ANOVA (ANalysis Of VAriance) is a process used to analyse a single variable's variance when gathered from multiple independent groups, each assumed to produce values which are normally distributed [13]. The approach taken compares local variance within groups to global mean variance between groups using an F-test. Firstly, a ratio is calculated between variance of means of each group and mean of within-group variance of each group, this is called the F-ratio. The F-ratio follows an F-distribution, interpreted as the probability (p-value) that the null-hypothesis is true for a given F-ratio, namely that each group produces values for the variable from the same normal distribution.

We use this p-value to order our physical traits on their ability to distinguish between individuals. We know our subjects are all distinct, subsequently we can define the most significant traits as those which refute the null hypothesis, minimise the p-value and therefore separate individuals most successfully.

Each of the 10 subjects had sample annotations from 38 annotators, resulting in 10x38 samples analysed per physical trait. The resulting p-values from ANOVA are shown in order in Fig. 2 using a logarithmic scale. **Race** and **Sex** features are the most significant by far; results expected from analysis of the literature and by choice of subjects. **Proportions** and **Leg Shape** had produced notable confusion amongst annotators during the annotation process and are also expectedly some of the least significant features. Both **Facial Hair** traits were the least significant, most probably since no subject had noticeable facial hair. It is of note that the six most potent traits would be very difficult to derive with any automated approach and are available by human description only.

##### B. ANOVA validation

To check whether the feature ordering recommended by ANOVA was meaningful, a *Leave-One-Out* classification test[14] was performed. A single annotation was left out (*the test set*) and compared to the remaining annotations (*the training set*). The training set was flattened between all annotators per subject, effectively generating a single feature vector representing each subject which was comprised of the mean annotations of all training set annotators. These training feature vectors were compared to the test annotation using a nearest neighbour classification scheme ( $k=1$ ) utilising the Euclidian distance between annotation feature

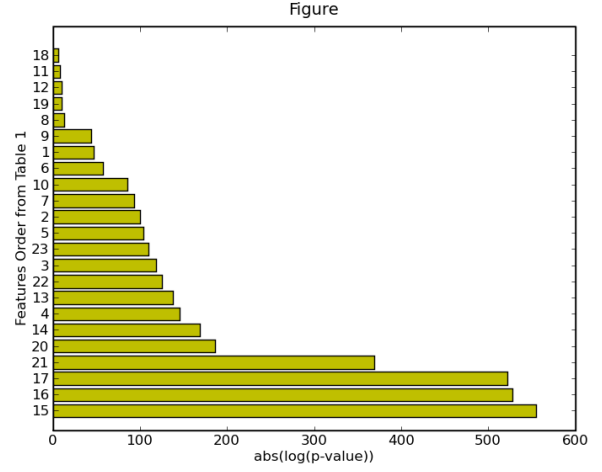


Fig. 2. Bargraph showing absolute log-p-values for each physical trait. A log scale is used to make differences visible regardless of large scale variations across p-values

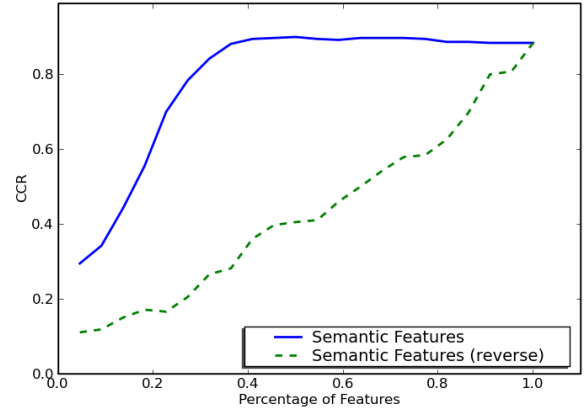


Fig. 3. CCR against number of features used. CCR calculated using kNN (with  $k=1$ ) and a take-one-out classification test. The graph compares the use of features in order of significance recommended by ANOVA and in the exact reverse order.

vectors. Each annotation collected was used as the test item exhaustively and a CCR was returned which was the ratio of correct classifications to number of classification attempts.

The CCR was measured using different combinations of features. We combined the features using two ordering methods: in the first we used the order suggested by the ANOVA ranking; in the second we reverse this order. The results show that using around 40% of the features (**Race** to **Weight**) achieves an optimal CCR of around 90% and also that the ANOVA ordering of features seems to be an improvement on the exact reverse of ANOVA; this gives us some confidence in this feature selection scheme.

##### C. Comparative CCR

To test the usefulness of semantic features in fusion as a soft biometric, two simple fusion strategies were implemented. To acquire an accurate CCR (Correct Classification

Rate) in both fusion scenarios, measures were taken to guarantee that no trivial comparisons were made, i.e. no 2 feature vectors were compared that contained identical sub components. Subsequently, rather than perform a fusion then take out feature vectors, the *Leave-One-Out* procedure took out annotators and subject sample pairs exhaustively; the fusion of which resulted in the test vector. The training set vectors fused the remaining samples (of every subject) with the remaining annotators. The annotation features used in each fusion scenario are the top 40% recommended by ANOVA where the number of automatic features used is changed from 0% to 100%, themselves ordered by ANOVA.

**Feature Fusion:** This approach assumes independence between the automatic and semantic data sources and concatenates the two feature domains. We also assume that the semantic annotations generated for a particular subject would have been generated identically regardless of sample video of that subject. This is reasonable as annotators were given access to all sample videos of a subject when making annotations. Therefore, each sample of each subject is repeated once per annotator, resulting in 38 fused feature vectors per subject sample, each having identical automatic signature components but distinct semantic components. Features themselves are normalised inherently due to construction, semantic labels are represented by numbers between 0 and 1 and Fourier component magnitudes are normalised to be between 0 and 1 during construction. Feature subset selection is also performed implicitly, with the high frequency components making up the gait signature and the top 40% semantic features making up the semantic signature. Classification is performed using kNN (k Nearest Neighbour) and the Euclidian distance between concatenated feature vectors.

**Score Fusion:** A transformation-based score fusion was implemented where the match scores generated by multiple biometric matchers are combined *directly* to create a new score [25]. The match score in our scenario is the *difference* between 2 sets of features (i.e. their Euclidian distance); the lowest score is taken as the closest match. The semantic annotations and visual signatures are our two distinct matchers. Each matcher attempts to find the best score of a given test sample when compared to *each subject* by comparing it to each training sample<sup>5</sup> of that subject. Scores for each subject are normalised using a min-max operation to account for scale differences between annotation and visual domains. The scores are then combined, for each subject, using a sum rule and the subject with the lowest score is taken as the estimated classification. The CCR is the ratio of correct classifications against total number of classification attempts.

#### D. Results

The results for these tests are shown in Fig. 4. The semantic traits in isolation achieved a maximum CCR of 90.0% where the automatic features in isolation achieved a maximum CCR of 98.1%. Upon closer inspection, the automatic feature CCR was the result of the miss classification of subject “023” (an average height Caucasian Male)

<sup>5</sup>either automatic signatures or annotations

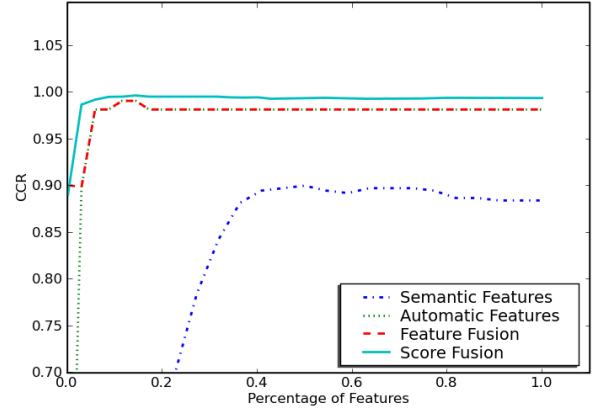


Fig. 4. The comparative CCR of the visual features against the semantic features and the 2 fusion schemes

as subject “083” (a relatively tall Chinese Male). Although it is conceivable that these individuals would produce similar automatic symmetry signatures, the production of similar semantic annotations is less likely. This is reflected in the fusion results both producing CCRs that improve upon the results of automatic features in isolation. Feature fusion achieves a marginal, though consistent improvement with a CCR of 98.2%, while a greater improvement is produced using score fusion which achieves a CCR of 99.5%. A possible reason for the gap could be the normalisation schemes used. In score fusion min-max normalisation was performed on the *match scores* obtained for each class separately for each matcher, where in feature fusion min-max normalisation was performed on each *feature* in isolation. This subtle distinction meant that in feature fusion the annotation features effectively had a smaller impact on overall score, having only 23 annotation features compared to 709 automatic features. This would also explain the similarity between feature fused results and automatic features alone.

#### V. CONCLUSIONS AND FURTHER WORK

We have introduced the use of semantic human descriptions as a soft biometric. We carefully selected a set of physical traits and successfully used them to annotate a set of subjects. Using ANOVA we outlined the most important semantic traits, results confirming prominent traits of previous studies. We have also shown that semantic traits have inherent identification capability, and also that they can successfully improve identification results of a primary gait biometric when combined in fusion.

In future work we hope to further explore the semantic space. We aim to find correlations that semantic features have with each other as well as correlations they share with automatic features. Discovery of these correlations will allow for the automatic generation of semantic annotations as well as the facilitation of content based searches using semantic queries.

We also hope to expand the corpus of semantic traits identified thus far to include other traits defining other



physical appearances, but also explore the annotation of subject actions and environments. Physical appearance traits such as clothing, piercing or distinguishing marks have yet to be explored but are often mentioned by witnesses of crime and help forge an annotators perception of a subject. Action descriptions, including description of gait (e.g. shuffle, limp, run), perceived mood, subject goals and social roles are often noticed according to the literature and complement dynamic aspects of gait, rather than the static aspects studied thus far. Finally, a subject's location and environment undoubtedly affect perception of subject features, but also define the concept of outliers and "Unusual" behaviour. Questions such as "Is this subject acting inappropriately?" or "Does this person look out of place?" are inherently related to the environments within which the subject is observed.

An exploration into these semantic attributes, supported by this initial work, will facilitate the involvement of human knowledge in biometric systems and also help bridge the semantic gap.

## VI. ACKNOWLEDGMENTS

The research was part sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon

## REFERENCES

- [1] G. Barbujani. Human races: Classifying people vs understanding diversity. *Current Genomics*, 6:215–226(12), June 2005.
- [2] A. Bertillon. *Signaletic Instructions including the theory and practice of Anthropometrical Identification*. The Werner Company, 1896.
- [3] V. Bruce and A. Young. Understanding face recognition. *British J. of Psych.*, 77:305–327, 1986.
- [4] G. B. Chapman and E. J. Johnson. *Incorporating the irrelevant: Anchors in judgments of belief and value*, pages 120–138. Heuristics and Biases: The Psychology of Intuitive Judgment. Cambridge University Press, 2002.
- [5] R. M. Dawes. Suppose We Measured Height With Rating Scales Instead of Rulers. *App. Psych. Meas.*, 1(2):267–273, 1977.
- [6] H. D. Ellis. *Practical aspects of facial memory*, section 2, pages 12–37. Eyewitness Testimony: Psychological perspectives. Cambridge University Press, 1984.
- [7] R. H. Flin and J. W. Shepherd. Tall stories: Eyewitnesses' ability to estimate height and weight characteristics. *Human Learning*, 5, 1986.
- [8] S. J. Gould. The Geometer of Race. *Discover*, pages 65–69, 1994.
- [9] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter. Automatic Gait Recognition by Symmetry Analysis. *Pattern Recognition Letters*, 24 (13):2175–2183, 2003.
- [10] D. J. Hurley, B. Arbab-Zavar, and M. S. Nixon. *The Ear as a Biometric*, chapter 7. Handbook of Biometrics. Springer, 2008.
- [11] Interpol. Disaster Victim Identification Form (Yellow). booklet, 2008.
- [12] A. K. Jain, A. Ross, and S. Prabhakar. An Introduction to Biometric Recognition. *Trans. CSVT*, 14:4–19, 2004.
- [13] R. A. Johnson and D. W. Wichern, editors. *Applied multivariate statistical analysis*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988. ISBN 0-130-41146-9.
- [14] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI*, pages 1137–1145, 1995.
- [15] P. V. Koppen and S. K. Lochun. Portraying perpetrators; the validity of offender descriptions by witnesses. *Law and Human Behavior*, 21 (6):662–685, 1997.
- [16] L. T. Kozlowski and J. E. Cutting. Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics*, 21: 575–580, 1977.
- [17] R. Lindsay, R. Martin, and L. Webber. Default values in eyewitness descriptions. *Law and Human Behavior*, 18(5):527–541, 1994.
- [18] M. D. MacLeod, J. N. Frowley, and J. W. Shepherd. *Whole body information: Its relevance to eyewitnesses*, chapter 6. Adult Eyewitness Testimony. Cambridge University Press, 1994.
- [19] C. N. Macrae and G. V. Bodenhausen. Social Cognition: Thinking Categorically about Others. *Ann. Review of Psych.*, 51(1):93–120, 2000.
- [20] J. M. Montepare and L. Zebrowitz-McArthur. Impressions of People Created by Age-Related Qualities of Their Gaits. *Journal of Personality and Social Psychology*, 55(4):547–556, October 1988.
- [21] K. Nandakumar, S. C. Dass, and A. K. Jain. Soft biometric traits for personal recognition systems. In *ICBA04*, pages 731–738, 2004.
- [22] M. Nixon and J. Carter. Automatic recognition by gait. *Proc. of the IEEE*, 94(11):2013–2024, Nov. 2006.
- [23] A. J. O'Toole. Psychological and Neural Perspectives on Human Face Recognition. In *Handbook of Face Recognition*. Springer-Verlag, 2004.
- [24] J. G. Ponterotto and B. Mallinckrodt. Introduction to the special section on racial and ethnic identity in counseling psychology: Conceptual and methodological challenges and proposed solutions. *J. of Counselling Psych.*, 54(3):219–223, July 2007.
- [25] A. Ross, K. Nandakumar, and A. K. Jain. *Introduction to Multibiometrics*, chapter 14. Handbook of Biometrics. Springer, 2008.
- [26] M. Savvides, J. Heo, and S. W. Park. *Face Recognition*, chapter 3. Handbook of Biometrics. Springer, 2008.
- [27] J. Shutler, M. Grant, M. S. Nixon, and J. N. Carter. On a large sequence-based human gait database. In *RASC '06*, pages 66–72, 2002.
- [28] S. V. Stevenage, M. S. Nixon, and K. Vince. Visual analysis of gait as a cue to identity. *App. Cog. Psych.*, 13(6):513–526, 1999.
- [29] H. Tajfel. Social Psychology of Intergroup Relations. *Ann. Rev. of Psych.*, 33:1–39, 1982.
- [30] N. F. Troje, J. Sadr, and K. Nakayama. Axes vs averages: High-level representations of dynamic point-light forms. *Vis. Cog.*, 14:119–122, 2006.
- [31] J. L. Wayman. Benchmarking Large-Scale Biometric System: Issues and Feasibility. In *CTST Government*, 1997.
- [32] G. L. Wells and E. A. Olson. Eyewitness testimony. *Ann. Rev. of Psych.*, 54:277–295, 2003.
- [33] A. D. Yarmey and M. J. Yarmey. Eyewitness recall and duration estimates in field settings. *J. of App. Soc. Psych.*, 27(4):330–344, 1997.
- [34] R. Zewail, A. Elsafi, M. Saeb, and N. Hamdy. Soft and hard biometrics fusion for improved identity verification. *MWSCAS*, 1:225–8, 2004.