Soft Biometric Fusion for Subject Recognition at a Distance

Bingchen H Guo, Mark S Nixon, and John N Carter

School of Electronics and Computer Science,
University of Southampton,
Southampton, United Kingdom
bingchenguo@yahoo.com, {msn,jnc}@ecs.soton.ac.uk

Abstract—There is societal need for techniques to identify subjects at a distance and when conventional biometrics are obscured, for example in fighting crime. Soft biometrics have this capability and include a subject's height, weight, skin colour and gender. Although the distinctiveness of soft biometric features is intuitively less than that of traditional biometric features, numerous experiments have demonstrated that the desired recognition accuracy can be achieved by using multiple soft biometric features. This paper will propose state-of-the-art multimodal biometric fusion techniques to improve recognition performance of soft biometrics.

The key contribution of this paper is the analysis of the influence of distance on soft biometric traits and an exploration of the potency of recognition using fusion at varying distances. A new soft biometric database, containing images of the human face, body and clothing taken at three different distances, was created and used to obtain face, body and clothing attributes. This new database was constructed to explore the suitability of each modality at a distance: intuitively, the face is suitable for near field identification, and the body becomes the optimal choice when the subject is further away. The new dataset is used to explore the potential of face, body and clothing for human recognition using fusion. We present a novel fusion technique at score and rank level that improves identification performance.

A novel joint density distribution-based rank-score fusion is also proposed to combine rank and score information. Analysis using the new soft biometric database demonstrates that recognition performance is significantly improved by using the new methods over single modalities at different distances.

Keywords— Soft biometric; recognition at a distance; joint density distribution; rank-score fusion

I. INTRODUCTION

SOFT biometrics are traditionally used to strengthen 'hard' biometric signatures: using, say, gender and age to improve performance [1]. Soft biometrics for identification concerns recognition using semantic descriptions as discriminatory features to identify subjects [2]. Here, we shall concentrate on those attributes used for identification. These approaches use computer vision/ machine learning and there is parallel interest in psychology concerned with identification, showing how trained (human) examiners have yet to use external face and body in recognition [4]. Soft biometric attributes include height, weight, gender and skin colour, which can be used to identify a person. In contrast, traditional biometrics overwhelmingly rely on sophisticated data collection devices. For example, facial recognition generally requires high image quality; however, image quality dramatically decreases as distance increases. Soft facial features, such as skin colour and face size, are relatively straightforward to perceive, even at a long distance. In comparison with traditional biometrics, soft biometric attributes are more easily understood.

Recognition performance using individual soft biometric datasets has been studied in previous research. A model using the human semantic description of soft biometrics to identify subjects was proposed in [5], where soft biometric features were used to enrich the recognition method. There were 19 body features investigated in [6] and the results demonstrated that shoulder shape and arm length can aid recognition. Another paper discussed 24 soft facial attributes [7], and their performance is measured through analysis of variance, entropy and mutual information. Skin colour, eyebrow length and face length were demonstrated to be more reliable for use in recognition. In addition, 21 clothing features were reported in [8], demonstrating that clothing features can also be used for recognition (though clothing is innately short term as clothes can easily be changed). Furthermore, it was also demonstrated that head coverage, lower body clothing category and belt presence can greatly improve recognition.

Most traditional biometric features distinguish people by using their distinctive features, such as DNA and fingerprints, whilst soft biometric features are not so discriminative by their nature. Nonetheless, accurate recognition can be achieved by using multi-modal soft biometrics. Despite the research into multi-modal soft biometric recognition being at an early stage, some articles have reported results using the most advanced methods. A method proposed in [9] used soft biometric attributes to improve recognition performance of traditional biometric. Prior work on soft biometrics at a distance is reported in [10], which demonstrated that the fusion of soft

biometrics and traditional facial features could improve the performance of recognition based on a sparse representation. A fusion method proposed in [11] used soft biometrics (body, clothing and face) for identification at a distance. The results demonstrated that Bayesian fusion can greatly improve recognition performance.

In order to achieve more accurate recognition, recognition systems frequently employ multi-modal fusion. Fusion approaches are conventionally divided into five different levels: sensor, feature, score, rank and decision [12].

Feature level fusion based on feature extraction from multiple data sources is intended to create a new feature set to represent a subject. Therefore, the key requirement of the technology is to effectively describe feature information in order to achieve the most accurate recognition. The general idea is to minimise the distance of feature information between intra-class samples, and maximise the distance between inter-class samples. Another important research field in the area of feature fusion focused on how to extract effective information by removing redundancy. Among a number of potential techniques used for feature extraction, linear feature extraction methods are widely used to reduce the dimensionality of the feature set. For example, a feature fusion method based on canonical correlation analysis (CCA) is introduced in [13]. Another feature level fusion technique, discriminant correlation analysis (DCA) [14], develops CCA by incorporating the class information into the correlation analysis of the feature sets. A multi-modal method, based on sparse representation, is proposed by Sumit, which significantly improved robustness and accuracy [15].

Score level fusion uses a combination of match scores from different biometric matchers, and then derives a new score from this information. Some simple methods, such as product rule, sum rule, max, medium and min rules, were introduced [16]. These methods can be readily implemented, since they do not require statistical information. Some score level fusion methods are based on the match score density distribution. A combinational method, using the Bayesian approach, was proposed in [16]. It estimated the genuine and impostor matching scores for each component modality. A support vector machine (SVM) based score level fusion was introduced and validated in [17]. The weighted score level fusion achieved a higher accuracy with the lower equal error rate (EER), compared with individual modalities [18].

A new rank and score level fusion method is presented here, based on joint density distribution. Since rank is a linear description (i.e. 1, 2, 3, …), it can be used to indicate the order of enrolled samples, but fails to describe the variations between adjacent samples. Thus, a novel technique is proposed to combine the effective information in rank and similarity scores, namely joint density-based rank-score fusion to consolidate the recognition result. The fusion effects over soft biometric characteristics and over distances are also analysed.

II. SOFT BIOMETRIC DATASET

For soft biometric recognition, there is no standardised dataset to evaluate recognition performance, and this is especially problematic in the research area of soft biometrics at different distances. One significant advantage of soft biometrics is that it does not have strict requirements on the resolution of collected images. The soft biometrics labels can be collected at a distance. Research into the influence of distance on feature annotation can give more useful information about features to be used at far or close distances. A new soft biometric database, based on different distances, was built to investigate this. In order to approach real life identification, the images in this database were simulated to be in an outdoor environment. Compared with the background controlled in a laboratory, there are more objects, such as buildings and cars, in the outdoor background, which can be used as points of reference when the soft traits are labelled. The accuracy of feature annotation can be improved with the help of reference objects.

A. Synthesising images

The new database, comprising of 131 male and 69 female subjects, was built and labelled with face, body and clothing traits. The original images were collected from the gait tunnel in University of Southampton [19], and then synthesised with an outdoor environment. The constant synthetic outdoor background removes any influence of change in indoor laboratory background, and avoids any change in illumination which is often experienced in outdoor imagery. There were 12 cameras deployed in different positions in the gait laboratory. The resolution of cameras is 640×480 and the capture rate is 30 frames per second. The viewpoint, at which the images collected provide the maximum face and body information of subjects was selected, as shown in Figure 1. The length of the gait laboratory is 7 meters. The minimum distance for the whole-body observation is 2 meters away from the camera so image acquisition occurs between 2 and 7 meters. Three points (2, 4.5 and 7 meters away from the camera) were marked as close, medium and far, respectively. Three images were selected in which the subjects stood at those three distances, and then simulated in an outdoor environment.

The outdoor environment required a geometry similar to that of the laboratory and the outdoor camera was positioned
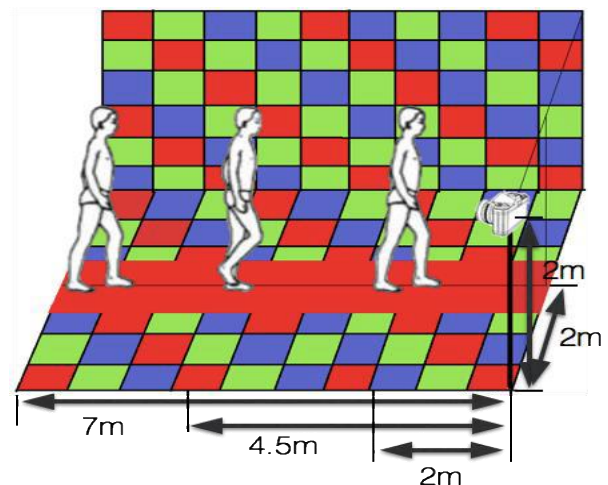

Fig. 1. Diagram of data acquisition environment

(a) Laboratory – close   (b) Subject extraction
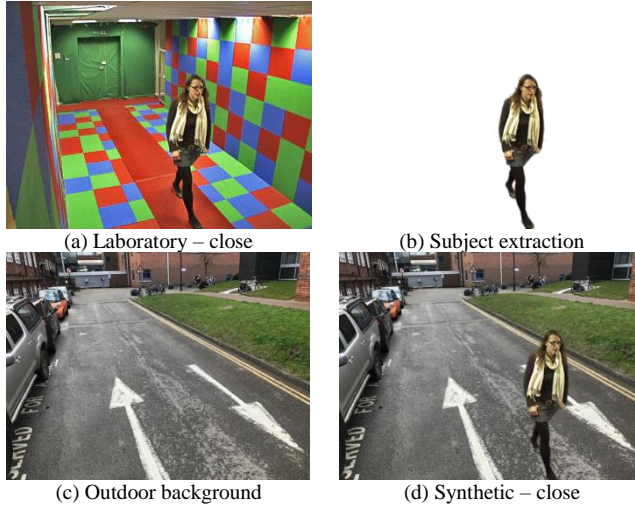
(c) Outdoor background   (d) Synthetic – close

Fig. 2. Synthesising images with outdoor environment

to be the same as that used in the laboratory. Three points at the close, medium and far distance were marked in advance. A cloudy day was chosen to eliminate shadows in the outdoor environment. Moreover, the brightness and contrast of laboratory images were improved to resemble outdoor conditions. In that way the outdoor images had a consistent background for all subjects at the three distances, and so the background could not affect recognition.

The next step was to segment subjects from laboratory images and to place them at appropriate locations in the outdoor background images. First, the subject's image is extracted from the laboratory image (Figure 2 (a), (b)) and then added to the background (Figure 2 (c)). A synthesised image is shown in Figure 2 (d). This process is applied to laboratory images acquired at the close, medium or far distances from the camera.

B. Soft biometric attributes and labels

The next step was to select soft biometric features for the body, face and the clothes that could be observed and described precisely and conventionally at different distances (i.e. skin colour and height). The facial features (i.e. the shape of the eyebrows and the length of the face) have strong discriminatory power; the body and clothing traits such as gender and the majority colour of clothes, more easily can be discerned. The earlier research studied recognition capability, and this part will utilise that work. It is possible to prune the feature set according to their recognition capabilities.

There were 19 body attributes analysed in [6]. The recognition performance of 12 body attributes was investigated in [20]. Considering the results in [6] and [20] , 10 effective attributes were selected as the body attributes for the new dataset. The attributes and labels are listed in Table I.

24 facial attributes were compared in [7]. ANOVA (Analysis of variance), Entropy and mutual information were employed to analyse the recognition performance of the attributes. The results demonstrated that Skin Colour, Eyebrow Length, Lip Thickness and Face Length have better recognition performance and are the most consistent. The

TABLE I
BODY TRAITS AND LABELS USED TO COMPARE SUBJECTS

Body traits	Labels
Gender	More feminine, Same, More masculine
Age	Older, Same, Younger
Height	Taller, Same, Shorter
Weight	Fatter, Same, Thinner
Shoulder shape	More square, Same, Rounder
Hair colour	Lighter, Same, Darker
Hair length	Shorter, Same, Longer
Neck length	Shorter, Same, Longer
Humpback	More straight, Same, More curved
Arm length	Longer, Same, Shorter

TABLE II
FACE TRAITS AND LABELS USED TO COMPARE SUBJECTS

Face traits	Labels
Eyebrow shape	More straight, Same, More curved
Nose shape	More flatter, Same, More protruding
Forehead	Straighter hairline, Same, More receded hairline
Eyes	Smaller, Same, Larger
Ears	More hidden, Same, More evident
Skin colour	Lighter, Same, Darker
Face size	Shorter, Same, Longer
Face	More bony, Same, Fleshier
Lips	Thinner, Same, Thicker
Chin and jaw	More angular, Same, Rounder

TABLE III
CLOTHING TRAITS AND LABELS USED TO COMPARE SUBJECTS

Clothing traits	Labels
Upper body clothing category	Jumper, T-shirt, Shirt, Blouse, Sweater, Coat, Hoodie, Other
Lower body clothing category	Trousers, Skirt, Dress
Any attached object category	None, Bag, Gloves, Hat, Scarf, Necktie, Other
Clothing style	Well-dressed, Business, Sporty, Fashionable, Casual, Other
The majority colour of upper body	Grey, Black, White, Jeans blue, Others
The majority colour of lower body	Grey, Black, White, Jeans blue, Others
Face coverage	Yes, No
head coverage	Yes, No
Presence of belt	Yes, No, Unsure
Wear glasses	Yes, No

facial traits were selected based on [7]. In this paper, the face images were collected at a close distance with high-quality images. Some features which cannot be observed from far away, such as Eye-to-Eyebrow Distance and Inter Eyebrow Distance, are modified in the new dataset. For example, Eyebrow Thickness and Eyebrow Length are replaced with Eyebrow Shape. The attributes and labels are listed in Table II.

In [8], 21 categorical traits and 7 comparative traits of clothing are analysed. The results demonstrated that clothing attributes can achieve good results when they are used for recognition. Furthermore, the accuracy using categorical traits is better than comparative traits. ANOVA was used to analyse the performance of different traits, and the results show that head coverage, lower body clothing category and belt presence are better identifiers than other traits. Thus, categorical traits

are used for the new clothing dataset. Seven features are investigated in [8], which have good recognition performance and are straightforward to observe at different distances, plus three new attributes: the majority colour of upper body and of lower body, and the presence of glasses, constitute the new clothing feature set. The attributes and labels for clothing are listed in Table III.

C. Data acquisition via Crowdsourcing

The soft features were labelled by human operators via crowdsourcing. A crowdsourcing task needed to be built for the large collection of high-quality comparative annotations. The CrowdFlower platform was used to build and run the crowdsourced annotation task. CrowdFlower provides comprehensive data analysis and quality control tools, allowing acceptance of a range of responses, whilst rejecting non-genuine answers.

Human categorical descriptions were originally acquired [2] based on subjective measurements that can vary between people due to differences in perception. Comparative descriptions [3] use features that are easily understood and annotated. Essentially by comparing features and then using a ranking algorithm each feature is projected along an axis which leads to more reliable measures. For height, as we shall find, subjects are not labelled as short or tall but the process provides a measures that indicates a subject's height relative to the population. The information for identification needs multiple comparisons between objects. Comparative descriptions can deliver more accurate descriptions. In this case, each of the new images was labelled with all 30 soft biometric traits. In data collection, categorical descriptions of clothing and comparative descriptions of the face and body

(a) Body labels

(b) Clothing labels

Fig. 3. Interface of label collection system

traits were used. It was demonstrated that the clothing traits could be used for recognition, and it is properly described using categorical labels [21]. In terms of face and body traits, comparative descriptions can convey more accurate descriptions, since observers can easily perceive differences between two subjects, for example, one person being taller than the other [6] [22]. This eliminates known psychological effects, such as owner variables and confirmation bias.

Each comparison describes the difference of each feature between two subjects, such as height, weight and the length of an arm. The comparison for each feature is labelled using three classes: shorter, the same or higher, following the observation that a scale of 3 could lead to better discriminative capability [7]. Each level is denoted by a signed integer, for example, when comparing the height of two subjects, -1 means shorter, 0 represents the same and +1 means taller. The labels for the new database based on the traits were collected using CrowdFlower. The interface for the collection system is shown Figure 3. Each of the 200 individuals was labelled by 20 people for the categorical clothing labels. The face and body were labelled by the comparison between each of the 200 subjects and 20 randomly chosen subjects. The total number of comparisons is 4000, each one labelled by 20 people.

III. RANK-SCORE FUSION

The results presented in previous research demonstrated the effectiveness of fusion in rank and score level using biometrics features for subject identification. The methods demonstrated that recognition performance can be improved by rank or score fusion. In this paper, an improved fusion method based on rank and score level fusion will be proposed. The difference between testing samples and enrolled samples can be observed intuitively using similarity scores, based on which ranks are sorted. However, the rank and similarity score information is different. Since rank is a linear description it can be used to indicate the order of enrolled samples, but does not describe the variation between adjacent samples. A novel technique will, therefore, be proposed to combine the effective information in rank and similarity scores, namely rank-score fusion, and then used to consolidate the recognition results.

A. Rank-score distribution

The outputs of a classifier always comprise a match score list or a rank list. The match score describes a distance or a similarity between the testing subject and the registered subject. The distance between subjects is calculated by Euclidean distance, Mahalanobis Distance or other metrics. The rank list is then obtained by sorting the similarity score in descending order.

Given k samples as registered samples for each matcher, a rank-score distribution can be constructed after training. In order to reduce the influence of outliers on the experiment and achieve more reliable results, each subject is used to match all remaining subjects. During the training process, similarity score matrix $s_i = \{s_{i,1}, \dots, s_{i,k-1}\}$, and rank matrix $r_i = \{r_{i,1}, \dots, r_{i,k-1}\}$ are obtained, where i is the subject ID in the
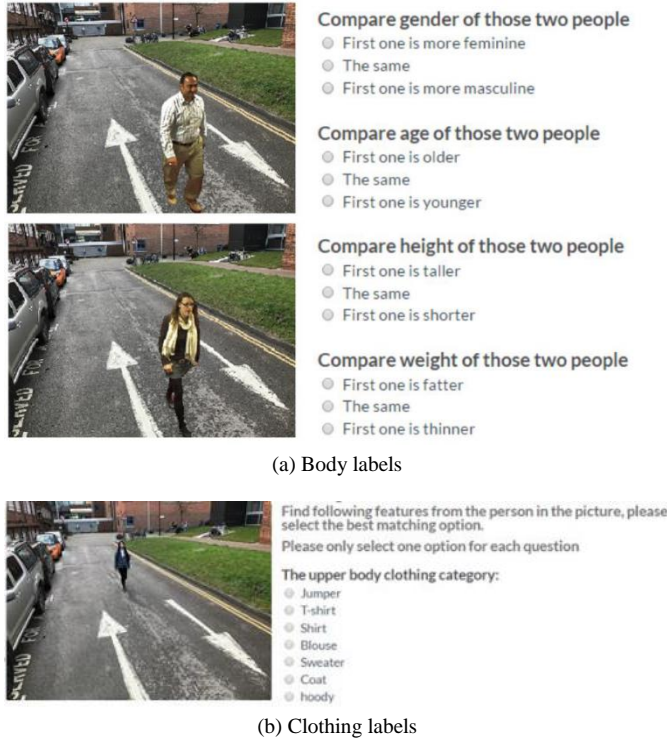
Fig. 4. Overview of the Rank-Score distribution calculation framework. The notation $sim(X_1, S_i)$ is used to denote the similarity score obtained by comparing an unknown subject S_i to the biometric sample 1 of a gallery subject in the face dataset.

range of 1 to $k - 1$, and $s_{i,n}$ is the similarity score between i^{th} subject with n^{th} remaining subject. The variables in the rank and score matrices are in one-to-one correspondence ($r_{i,n}$ is the rank order of $s_{i,n}$). For multi-modal biometrics, there are m different matchers, each of which produces one rank matrix and one score matrix. $S = \{[s_{i,1}^j, \dots, s_{i,k-1}^j]\}$ and $\mathcal{R} = \{[r_{i,1}^j, \dots, r_{i,k-1}^j]\}$ are used to denote the sets of scores and rank matrices separately, where j is the number of biometric matcher ($j = 1, \dots, m$) and $i = 1, \dots, k$.

As mentioned in the last paragraph of introduction, both score and rank fusion have their own advantages and limitations. Our rank-score fusion method is proposed here to combine the complementary information in rank and similarity scores. A two-dimensional density distribution with two corresponding variables in S and $\mathcal{R}$ is estimated. The calculation is constructed using a Gaussian Kernel function to smooth the result. The density function with rank and score is estimated by:

$$p(s, r) = \frac{1}{2\pi\sigma_r\sigma_s} exp\left[-\left(\frac{(r-\mu_r)^2}{\sigma_r} + \frac{(s-\mu_s)^2}{\sigma_s}\right)\right] \quad (1)$$

B. Normalization

After $p(s, r)$ is obtained, the joint density of a pair of match score and rank, $p(s_m, r_m)$, is calculated and employed as measurement of weights. In order to further improve the method, the density is normalised before it is used as a weight.

Z-score normalization: This method requires the arithmetic mean and standard deviation of a given score list. The normalisation is:

$$s_k' = \frac{s_k - \mu}{\sigma} \quad (2)$$

where s_k' is an updated score, μ is the arithmetic mean and σ is the standard deviation.

Min-max normalisation: Min-max normalisation is suitable for the case that the boundaries (minimum and maximum values) of the scores are known. After normalisation, all scores are in the range [0,1]. Given a set of match scores $S = \{s_1, \dots, s_k\}$, the normalisation scores are given by:

$$s_k' = \frac{s_k - min}{max - min} \quad (3)$$

Medium and median absolute deviation normalisation (MAD): Compared with Z-score normalisation, MAD is a robust method, since it is insensitive to outliers. The scheme of median and MAD is given by:

$$s_k' = \frac{s_k - med}{MAD} \quad (4)$$

where $med = median\{s_1, \dots, s_k\}$ and $MAD = median\{|s_k - med|\}$. This technique does not retain input distribution and does not transfer scores in a common range.

Tanh-estimators: The tanh-estimator proposed in [23] is a robust and efficient method. The normalised scores are given by:

$$s_k' = \frac{1}{2}\left\{tanh\left(0.01\left(\frac{s_k - \mu}{\sigma}\right)\right) + 1\right\} \quad (5)$$

where μ is the arithmetic mean and σ is standard deviation.

The recognition performance of the rank-score fusion method using different normalised algorithms is shown in Table IV. Tanh-estimators are demonstrated to provide the best result across different normalisation techniques. At medium and far distances, the results given by tanh-estimators has a distinct advantage, and is as good as MAD at a close distance.

TABLE IV
ACCURACY USING DIFFERENT NORMALISATION METHODS

	Close	Medium	Far
Non-normalised	96.5%	84.9%	74.0%
Min-max	96.9%	85.7%	77.0%
Z-score	96.7%	90.8%	78.8%
MAD	98.6%	86.9%	74.8%
Tanh-estimators	98.5%	92.5%	82.6%

C. Rank-score fusion

The parameters of the rank-score distribution specified in Eq.(1) are calculated from training data before testing. During testing, an unknown user is matched with all registered users. For each matcher, one similarity score list and one rank list is obtained. Each registered user has a corresponding similarity score and a rank order, which are used as the inputs in Eq.(1). The joint density of similarity score and rank order is calculated. After normalisation, it is used as a weight to update the similarity score. The final similarity score of an unknown
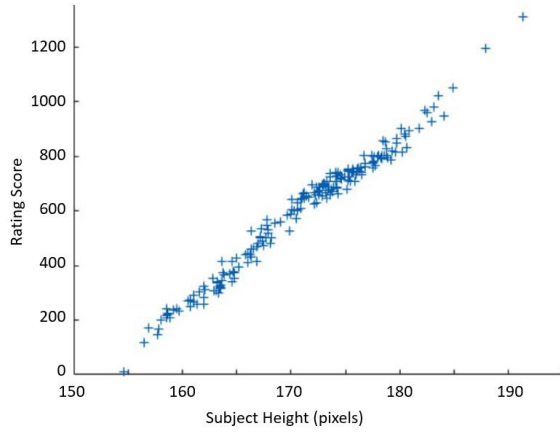
Fig. 5. Relationship between estimated and measured height

user and the kth enrolled user can be calculated using Eq.(6), which is a weighted sum of different matchers.

$$ws_s(k) = \sum_{j=1}^{m} w_{p_{j,k}} s_{j,k} \qquad (6)$$

where $s_{j,k}$ is the similarity score of enrolled user k and $w_{p_{j,k}}$ is normalised joint density. The unknown user is labelled as the class that has the maximum values of ws_s.

IV. EXPERIMENT AND DISCUSSION

A. Attributes Analysis

1) Ranking inference

The Elo rating system provides a ranking method based on Thurstone's case for comparative descriptions [24]. The Elo rating system was originally developed to quantify chess players' rankings. In this case, Elo was used to rank the comparative data. It is necessary to ensure the Elo rating results are trustworthy. Pixel height is used to estimate the actual height of a subject in an image. Meanwhile, an Elo ranking result, in terms of human height, represents a systematic judgement of height according to the comparative description. The positive correlation of pixel height and Elo rating results can validate the accuracy of Elo rating results, as shown in Figure 5.

The Pearson's correlation coefficient is 0.93 using this method, and outperforms the result 0.87 obtained previously [25]. This is likely due to the fixed geometry used from an outdoor environment, which allows labellers to pay more attention to the comparison of height.

2) Correlation analysis

The correlation coefficient (Pearson's) of each semantic feature in three groups (close and medium, close and far, medium and far) was employed to measure the stability of the ranking system. Theoretically, a larger coefficient indicates that a particular trait is less sensitive to distance.

The correlation coefficient of each trait in three groups is depicted in Figure 6. It shows that the stability of most of the clothing traits, and some of the body traits, is relatively high, whilst the stability of face traits varies substantially at different distances. For example, in the body feature set, humpback and
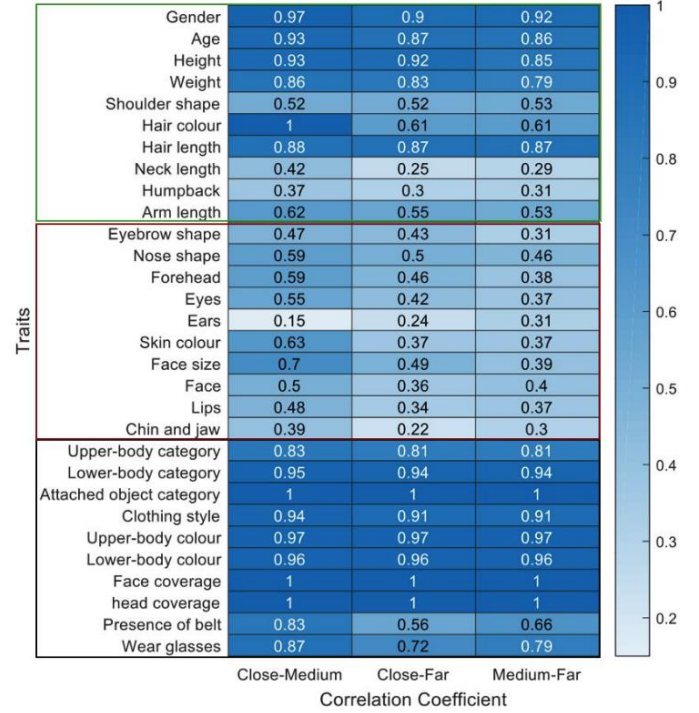

Fig. 6. Pearson's correlation coefficient for each trait in three groups (close and medium, close and far, medium and far)

neck length are the most sensitive traits. Hair colour is good between close and medium distance, but worse at close-far and medium-far. The result demonstrated that the hair colour is less stable at a far distance. In the facial feature set, the ear is weak at all three groups, which means it is a sensitive feature, whilst face size has the highest stability in the face trait set. In clothing traits, face coverage and head coverage are equal to one at all three groups. In other words, these two features are the most straightforward to be observed.

3) Mutual information

Mutual information was introduced to measure the strength of the dependency between two variables. Given two random variables, X and Y, whose marginal probability distribution functions are $p(x)$ and $p(y)$ respectively, the mutual information $I(X;Y)$ was given in

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) log \frac{p(x,y)}{p(x)p(y)} \qquad (7)$$

Here mutual information is used to measure the relevance between each trait and subject ID. Since subject ID presents the explicit differences of each subject, larger mutual information demonstrates the stronger discriminating capacity of the trait. Moreover, small differences of mutual information for each trait at three distances reflect superior reliability.

Figure 7 shows that age and upper body clothing category have the highest discriminatory differentiating power on body and clothing trait sets, respectively.

Despite skin colour showing the highest discriminatory power at close and medium distance, it is relatively weak at a far distance. It can be concluded that body traits have the highest discriminating power at different distances, whilst the clothing traits are more stable across all distances. It appears
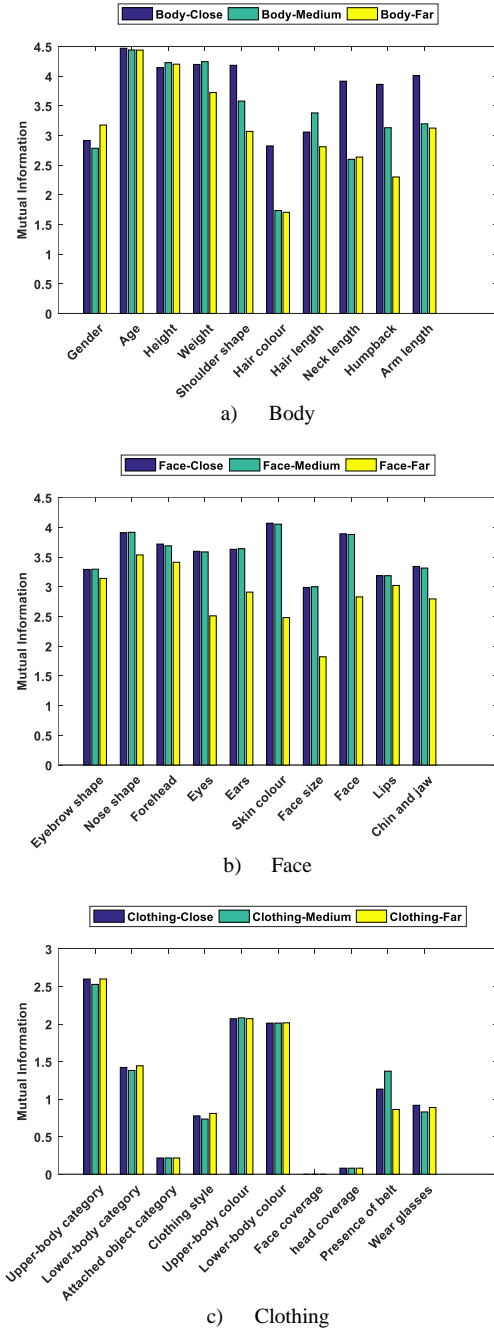
a)   Body

b)   Face

c)   Clothing

Fig. 7. Mutual information between trait and subject ID at three distances.

that clothing traits are more easily observed in comparison with the other two types of traits. However, they have less distinctiveness in recognition, which reduces discriminatory power. Further, clothing is an innately short-term biometric, since clothing can easily be changed. Moreover, since body traits have more detail than clothing traits, they have stronger differentiating power. It is known that the resolution of the face varies greatly with distance. Thus, the discriminatory power of facial traits decreases sharply at far distances. Furthermore, the trait stability evaluated by mutual information is in accordance with the results given by correlation analysis. The clothing traits appear to be the most stable traits for recognition.

B. Single-modal recognition

The purpose of the identification experiment is to assess the effectiveness of the proposed attributes (listed in Table I, Table II and Table III) for identification using the new dataset, and show the applicability of single-modal soft biometrics. The experiments simulate a realistic scenario that aims to retrieve the identity of an unknown subject (or probe) from a soft biometric database using verbal descriptions for the probe (i.e. eyewitness statement).

The experiments used LoO (Leave-one-out) cross-validation, and were implemented using 200 subjects, in which 100 subjects were selected at random as a gallery, and the remaining samples were used for testing.

The identification of unknown subjects was performed by k-nearest neighbors algorithm, the distance is calculated by the Euclidean distance d, between the biometric signature of the probe and the biometric signature of each subject in the gallery as follows:

$$d = \sqrt{\sum_{i=1}^{T}\big(X(i) - Y(i)\big)^2} \tag{8}$$

where X is a vector that represents the biometric signature of one subject. For example, the unknown subject, Y, is a vector that represents the biometric signature of another subject (the subject in the gallery that is compared with the unknown subject), and $T = 10$ is the number of soft biometric attributes composing the biometric signatures. The nearest neighbour was used here for classification: subjects were sorted in ascending order according to their distances from the probe, and the rank of the correct match was used to report the identification performance.

For face and body, the comparative labels are used. The recognition accuracy (rank=1) over varying numbers of probe comparisons (n) is shown in Figure 8 and Figure 9. It is easy to obtain that with increase in the numbers of comparisons; the recognition accuracy is improved accordingly. The similarity of Figure 8 and Figure 9 demonstrates that the number of comparisons directly influenced the recognition performance.

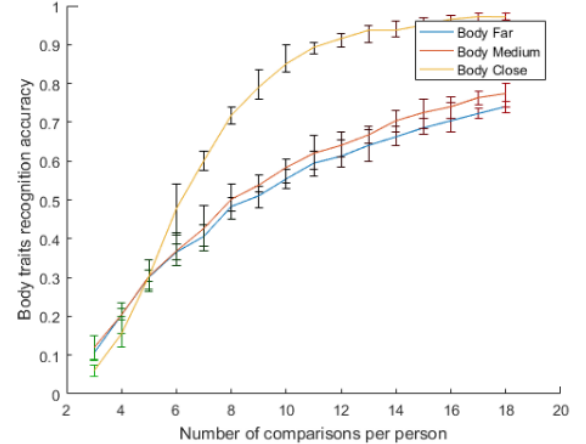Figure 8 shows body recognition accuracy obtained from

Fig. 8. Body recognition accuracy obtained from different numbers of comparisons.
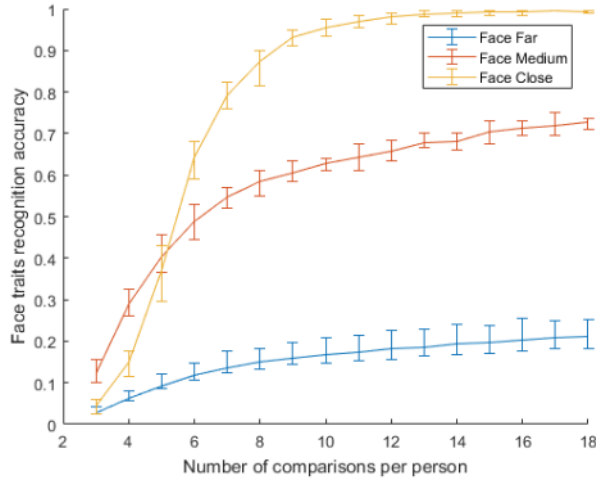
Fig. 9. Face recognition accuracy obtained from different numbers of comparisons.

Fig. 10. Cumulative match characteristic curves for the individual modalities

different numbers of comparisons. In this case, 9 comparisons are required for recognition. The recognition accuracy used one comparison only to achieve accuracy, of roughly 10% at all three distances. The recognition performance continues to increase over the range. At close distance, the accuracy improvement is noticeable when the number of people compared is lower than 12 and achieves ~98% with 18 comparisons. For medium and far distances, a ~85% correct recognition rate is achieved with 18 comparisons.

In comparison with [20], the database is comprised of 100 subjects and 12 body traits represented by comparative distributions. Their identification rate with 10 comparisons is around 97%, which is higher than that of the new method (around 85%). Nevertheless, the dataset used in this chapter is twice as large as the dataset in [20], and employs less features. This study concentrates on the fusion approaches and on whether fusion itself can be used for recognition as well as its properties, whilst assuming that all single mode approaches can be improved independently.

The facial recognition accuracy over different numbers of comparisons is shown in Figure 9. It shows that, at close distances, the accuracy of facial descriptions greatly outperforms that of body descriptions. The facial description achieves close to 80% identification accuracy with 7 comparisons, whilst body only achieves around 60%. The recognition performance at a close distance is much better than that at medium and far distances. A 97% recognition accuracy is obtained with ten comparisons, obtaining a maximum of a 100% accuracy at 18 comparisons. The recognition performance is limited at a far distance, with the accuracy rate achieving only 19% when using 18 comparisons. This suggests that distance is more important than the number of comparisons when using the face for recognition.

By comparing the identification performance at close distance obtained from [6], which achieved an accuracy of 100% with 10 comparisons by using 24 attributes with 4038 subjects, the accuracy of the proposed method using 10 comparisons is ~97%, which demonstrates that 10 traits used here include enough information for classification.
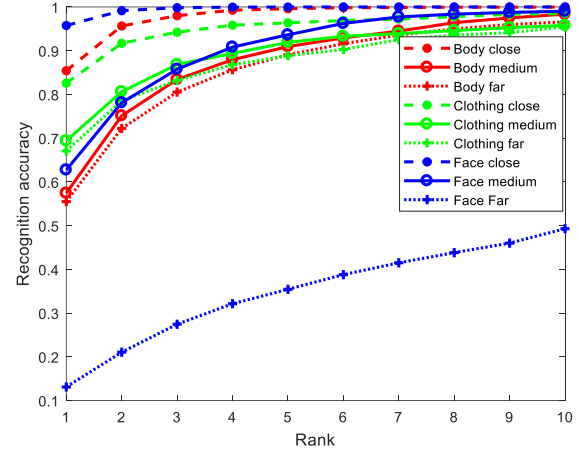
The cumulative match characteristic curves (CMC) for single-modal soft biometric feature sets are depicted in Figure 10. At a close distance, the recognition accuracy of facial traits is ~96% % (for rank 1 identification), but it falls sharply to ~63% at a medium distance and ~13% for a far distance. The recognition by clothing is the most consistent, with accuracies of ~83% ~69% and ~67% at the three distances respectively.

For facial traits, the recognition accuracy at close distance is 99.2% (rank=2) and 100% (rank=6). At medium and far distances, the recognition accuracy is 99.03% (rank=10) and 49.3% (rank=10) respectively. In terms of recognition using clothing traits, the accuracy at medium (95.8%) and far (95.9%) distances is very similar. It performs higher at a close distance (98.5%). For the body traits, the recognition accuracy is 99.3% (rank=4) at a close distance and decreases slightly to 98.4% and 96.6% (rank=10) at medium and far distances. In summary, at a close distance, the facial traits, as well as body and clothing traits, show high performance. At medium distance, three feature sets obtain similar accuracy, but at a far
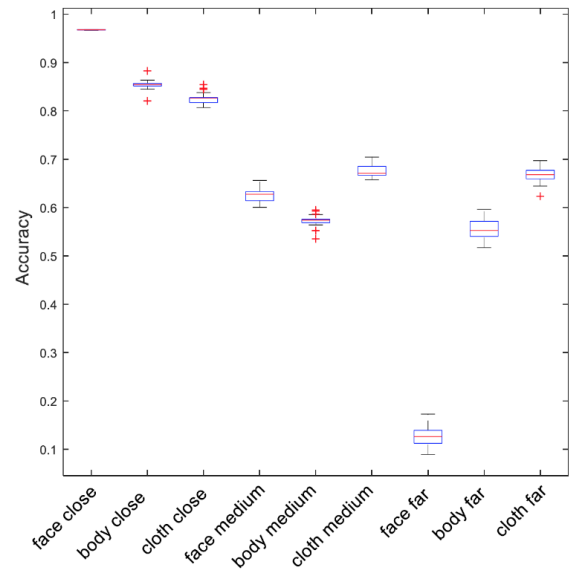

Fig. 11. Accuracy of single-modal recognition (rank=1).

TABLE V
IDENTIFICATION PERFORMANCE USING DIFFERENT FUSION METHODS

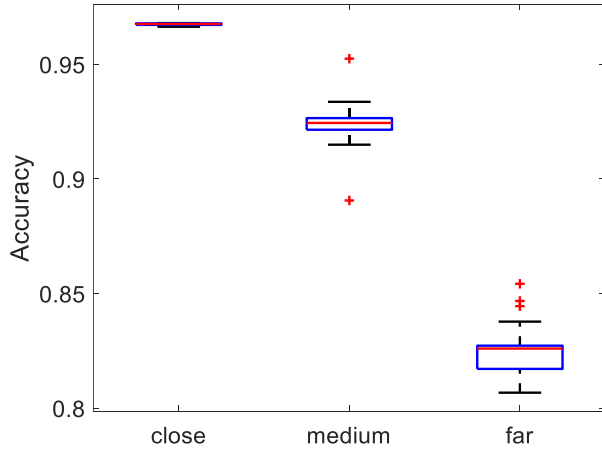|	Close		Medium		Far	
	Accuracy(%)	EER(%)	Accuracy(%)	EER(%)	Accuracy(%)	EER(%)
Bayesian theory [11]	96.3	0.38	84.6	1.07	78.1	2.57
Log likelihood ratio [26]	96.1	1.07	87.7	2.46	76.5	3.02
Logistic regression [27]	96.4	0.39	82.3	3.73	75.5	3.83
Nonlinear weight ranks [28]	96.9	0.39	86.2	3.48	79.3	3.44
PAV based [29]	97.0	0.38	86.0	3.01	79.1	3.33
Rank-score fusion	98.5	0.33	92.5	0.69	82.6	2.42

Fig. 12. Accuracy of fusion at three distances (rank=1).

distance, the accuracy of clothing and body traits significantly outperforms that of the face.

The process is repeated 20 times, and the boxplot of recognition results is shown Figure 11. It shows the extent of the accuracy over 9 single-modal methods. It shows that the recognition performance of clothing and body traits is relatively stable, whilst the performance of facial traits varies substantially at different distances. As we know, the resolution of face images is greatly influenced by distance. Thus, the recognition accuracy of facial traits decreases sharply at far distances.

C. Evaluation of rank-score fusion at three distances

In this section, experiments are performed to validate the performance of the proposed rank-score fusion algorithm at three distances.

The feature sets used are the same as those described in Section II. The experiments are implemented using 200 subjects. Each subject has 20 samples, in which 5 samples are randomly chosen to train the joint density function and to obtain normalisation weights. The remaining samples are used for testing. The experiments were repeated 20 times, and the box-plot of recognition results are shown in Figure 12.

It is clear that at the close distance is the most consistent, with an average recognition accuracy of 98.5% at close distance. The recognition accuracy at a medium distance

slightly decreases to 92.5% on average, with the maximum 95.3% and minimum 89.1%. The accuracy and the uncertainty at a far distance are worse than that at close and medium distances.

D. Compared with single-modal recognition

Figure 11 is a vertical boxplot that shows the extent of the accuracy over nine single-modal methods. At the close distance, the recognition accuracy of facial traits is 95.7%, which is the best modality when close. The average recognition accuracy is 98.5% after rank-score fusion, which is 2.8% higher than using single facial traits. The fusion result at a close distance is the most consistent, and the variance is smallest compared with the other two distances. The results demonstrate that clothing traits achieve the highest recognition rate at a medium distance (69.4%), which is increased by 23.1% using rank-score fusion. The stability of fusion results at a far distance is not as good as the other two distances, because the accuracy of facial traits at a far distance is only 13.1%, which lowers the fusion result. Meanwhile, the fusion results of rank-score fusion are slightly improved at a far distance, and the accuracy of the proposed rank-score fusion increases to 82.6%.

E. Comparison with other fusion methods

The recognition results at three distances using different fusion methods are listed in Table V. It is clear that the proposed method outperforms a selection of rank and score level fusion techniques. At all three distances, the recognition accuracy of the proposed fusion method is always superior to that of other methods. At a close distance, all the fusion methods achieve excellent recognition rates. Compared with the best results given by other methods (97.0%), rank-score fusion improves the accuracy by 1.5%. At a medium distance, the recognition performance of rank-score fusion improves significantly. The accuracy increases to 92.5%, and the EER achieves 0.69%. The improvement of fusion performance at close and far distances is not as apparent as at a medium distance, which improves recognition accuracy by 3.3% (Nonlinear weight ranks).

V. CONCLUSIONS

This paper introduces a new database for soft biometrics based on imagery collected from the University of

Southampton Gait Tunnel, synthesised so as to appear to be from an outdoor environment, and then labelled using CrowdFlower. The influence of distance on soft biometric traits was firstly analysed. In terms of single-modal recognition, facial traits achieve the best result at close distance but are not stable, they fall sharply with an increase in distance. Compared with facial traits, as clothing and body traits have less distinctiveness, the accuracies of the two trait sets are lower than that of facial traits at close distances, whereas the stability of body and clothing is much better than stability for the face. They can achieve good recognition results, even at a far distance.

A novel joint density-based rank-score fusion technique to fuse three soft biometric methods was proposed in this chapter. The experiments were conducted with other multi-modal fusion methods to make comparisons. Accuracy and EER were employed to evaluate their performance. The result of rank-score fusion demonstrates that at a close distance the soft biometric recognition performance is the most consistent. Compared with other fusion methods, the proposed rank-score fusion is numerically demonstrated to be able to obtain the best results at all three distances, particularly at a medium distance. This leads to more general conclusions on this work.

In conclusion, compared with single-mode biometric methods, the experimental results demonstrate clearly that the recognition performance of multi-modal soft biometrics is significantly improved by using biometric fusion. Naturally as recognition at a close distance is generally good, it can be seen that fusion largely improves the recognition at medium and far distances. There is still much room for improvement at the far distance. It would appear that the fusion process selects information that is best for recognition at any distance and so the effects are most dramatic when appropriate features are weighted more favourably for recognition purposes.

REFERENCES

[1] A. Dantcheva, P. Elia and A. Ross, "What else does your biometric data reveal? A survey on soft biometrics," IEEE Transactions on IFS, vol. 11, no. 3, pp. 441--467, 2015.

[2] M. S. Nixon, P. L. Correia, K. Nasrollahi, T. B. Moeslund, A. Hadid and M. Tistarelli, "On soft biometrics," Pattern Recognition Letters, vol. 68, pp. 218--230, 2015.

[3] Y. Hu, K. Jackson, A. Yates, D. White, P. J. Phillips and A. J. O'Toole, "Person recognition: Qualitative differences in how forensic face examiners and untrained people rely on the face versus the body for identification," Visual Cognition, vol. 25, no. 4-6, pp. 492--506, 2017.

[4] S. Samangooei, B. Guo and M. S. Nixon, "The use of semantic human description as a soft biometric," in Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on, IEEE, 2008, pp. 1--7.

[5] D. A. Reid, M. S. Nixon and S. V. Stevenage, "Soft biometrics; human identification using comparative descriptions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 6, pp. 1216--1228, 2014.

[6] N. Almudhahka, M. S. Nixon and J. Hare, "Semantic face signatures: Recognizing and retrieving faces by verbal descriptions," IEEE Trans. IFS, vol. 13, pp. 706-716, 2018.

[7] E. Jaha and M. S. Nixon, "Clothing Analysis for Subject Identification and Retrieval," in Recent Advances in Intelligent Image Search and Video Retrieval, Springer, 2017, pp. 167--211.

[8] A. K. Jain, S. C. Dass and K. Nandakumar, "Soft biometric traits for personal recognition systems," in Biometric Authentication, Springer, 2004, pp. 731--738.

[9] P. Tome, J. Fierrez, R. Vera-Rodriguez and M. S. Nixon, "Soft biometrics and their application in person recognition at a distance," IEEE Trans. IFS, vol. 9, no. 3, pp. 464--475, 2014.

[10] M. S. Nixon, B. H. Guo, S. V. Stevenage, E. S. Jaha, N. Almudhahka and D. Martinho-Corbishley, "Towards automated eyewitness descriptions: describing the face, body and clothing for recognition," Visual Cognition, vol. 25, no. 4-6, pp. 1--15, 2016.

[11] A. Ross and R. Govindarajan, "Feature level fusion of hand and face biometrics," in Biometric Technology for Human Identification II, vol. 5779, International Society for Optics and Photonics, 2005, pp. 196--204.

[12] Q.-S. Sun, S.-G. Zeng, Y. Liu, P.-A. Heng and D.-S. Xia, "A new method of feature fusion and its application in image recognition," Pattern Recognition, vol. 38, no. 12, pp. 2437--2448, 2005.

[13] M. Haghighat, M. Abdel-Mottaleb and W. Alhalabi, "Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition," IEEE Transactions on Information Forensics and Security, vol. 11, no. 9, pp. 1984--1996, 2016.

[14] S. Shekhar, V. M. Patel, N. M. Nasrabadi and R. Chellappa, "Joint sparse representation for robust multimodal biometrics recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 1, pp. 113--126, 2014.

[15] A. A. Ross, A. K. Jain and K. Nandakumar, "Score level fusion," in Handbook of Multibiometrics, Springer, 2006, pp. 91--142.

[16] F. Wang and J. Han, "Multimodal biometric authentication based on score level fusion using support vector machine," Opto-electronics review, vol. 17, no. 1, pp. 59--64, 2009.

[17] H. M. Sim, H. Asmuni, R. Hassan and R. M. Othman, "Multimodal biometrics: Weighted score level fusion based on non-ideal iris and face images," Expert Systems with Applications, vol. 41, no. 11, pp. 5390--5404, 2014.

[18] R. D. Seely, S. Samangooei, M. Lee, J. N. Carter and M. S. Nixon, "The University of Southampton multi-biometric tunnel and introducing a novel 3d gait dataset," in International Conference on BTAS, 2008.

[19] D. Martinho-Corbishley, M. S. Nixon and J. N. Carter, "Super-fine attributes with crowd prototyping," IEEE Transactions on Pattern Analysis and Machine Intelligence (Accept).

[20] E. Jaha and M. S. Nixon, "From Clothing to Identity: Manual and Automatic Soft Biometrics," IEEE Trans. on IFS, vol. 11, no. 10, pp. 2377--2390, 2016.

[21] S. Samangooei, B. Guo and M. S. Nixon, "The use of semantic human description as a soft biometric," in Biometrics: Theory, Applications and Systems, IEEE, 2008, pp. 1--7.

[22] A. Jain, K. Nandakumar and A. Ross, Score normalization in multimodal biometric systems, vol. 38, Pattern recognition, 2005, pp. 2270--2285.

[23] K. Tsukida and M. R. Gupta, " How to analyze paired comparison data," in WASHINGTON UNIV SEATTLE DEPT OF ELECTRICAL ENGINEERING, 2011.

[24] D. A. Reid and M. S. Nixon, "Using comparative human descriptions for soft biometrics," in Biometrics (IJCB), 2011 International Joint Conference on, IEEE, 2011, pp. 1--6.

[25] K. Nandakumar, Y. Chen, S. C. Dass and A. Jain, "Likelihood ratio-based biometric score fusion," IEEE transactions on PAMI, vol. 30, no. 2, pp. 342--347, 2008.

[26] A. A. Ross, K. Nandakumar and A. Jain, "Rank level fusion," in Handbook of multibiometrics, Springer Science & Business Media, 2006, pp. 70-72.

[27] A. Kumar and S. Shekhar, "Personal identification using multibiometrics rank-level fusion," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 41, no. 5, pp. 743--752, 2011.

[28] S. Nanang, "Pool Adjacent Violators Based Biometric Rank Level Fusion," in International Conference of the Biometrics Special Interest Group (BIOSIG), 2017.
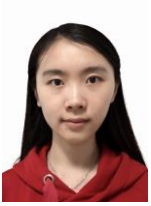
Bingchen H Guo received the MEng degree in Computer Science in 2014 and was working towards her Ph.D. in human identification with soft biometrics at the University of Southampton, U.K. Her research interests include computer vision, machine learning and visual perception with semantic image discrimination.

Mark S. Nixon is currently a Professor of Computer Vision with the University of Southampton, U.K. His research interests are in image processing and computer vision. His team were early workers in face recognition and later pioneers of gait recognition. He has chaired/program chaired many conferences (BMVC 98, AVBPA '03, FG '06, ICPR '04, ICB '09/15, and BTAS '10). He is a member of IAPR TC4 Biometrics, the IEEE Biometrics Council and Fellow of IET, IAPR, and BMVA.

John N. Carter (M'90) received the B.A. degree in experimental physics from Trinity College, Dublin, Ireland, and the Ph.D. degree in astrophysics from the University of Southampton, U.K. In 1985, he joined the School of Electronics and Computer Science as a Lecturer researching in signal and image processing, where he is currently a Senior Lecturer. His research interest is in the area of 4-D image processing, with applications in object tracking, feature detection, biometrics and automatic gait analysis.