Towards Automated Eyewitness Descriptions: Describing the Face, Body and Clothing for Recognition

Mark S Nixon[+], Bingchen H Guo[+], Sarah V Stevenage[*], Emad S Jaha[+], Nawaf Almudhahka[+] and Daniel Martinho-Corbishley[+]

[+]*Department of Electronics and Computer Science and* [*]*Department of Psychology, University of Southampton, Southampton, UK*

Corresponding Author

Professor Mark Nixon

School of Electronics and Computer Science University of Southampton, SO17 1BJ, UK

Tel: UK (0)23 8059 3542

msn@ecs.soton.ac.uk

Abstract

A fusion approach to person recognition is presented here outlining the automated recognition of targets from human descriptions of face, body and clothing. Three novel results are highlighted. First, the present work stresses the value of comparative descriptions (he is taller than…) over categorical descriptions (he is tall). Second, it stresses the primacy of the face over body and clothing cues for recognition. Third, the present work unequivocally demonstrates the benefit gained through the combination of cues: recognition from face, body and clothing taken together far outstrips recognition from any of the cues in isolation. Moreover, recognition from body and clothing taken together nearly equals the recognition possible from the face alone. These results are discussed with reference to the intelligent fusion of information within police investigations. However, they also signal a potential new era in which automated descriptions could be provided without the need for human witnesses at all.

1. Introduction

We live in a technologically sophisticated world in which the methods of police investigation are changing. There is a smartphone user, or a CCTV camera virtually on every street corner who can capture a perpetrator in the act. As such, imagery is often broadcast with the hope of eliciting public identification, and surveillance images can provide direct evidence of value to both the investigative process and the court system. Nevertheless, for all its sophistication, this information is useless unless we can make an identification from the images available. Two problems may prevent this. First, the images can sometimes be too poor in quality to enable fine-grained biometric analysis of characteristics. Second, the images may depict a perpetrator who purposely hides or disguises their face. The latter case was exemplified by the recent images of Jihadi John who hid his face (but not his body) in incriminating photographs. Consequently, a question arises as to whether the continuing focus on faces is appropriate if the face can so easily be degraded or hidden? The purpose of the present paper is to examine a new approach in computer vision which relies on soft biometrics. Specifically, we examine the utility of soft biometric descriptions of the face, body and clothing of a target when making an identification, and we explore the benefits that are possible when soft biometrics are combined in sensible and realistic ways.

1.1 Soft Biometrics for Identification

Soft biometrics represent a relatively new form of biometric identification which rely on the physical or behavioural characteristics as described by humans (Dantcheva, Elia & Ross, 2016; Nixon, Correia, Nashrollahi, Moeslund, Hadid & Tistarelli, 2015;). Earliest references described soft biometrics as descriptive labels which could be used to

separate populations into subsets (i.e., male, Caucasian) but which were not sufficient when trying to identify a specific individual. Later, soft biometrics were defined as the 'personal characteristics describable by humans that can be used to aid or effect person recognition' (Nixon et al., 2015, p220). Their value comes from the fact that they may help to refine a more traditional biometric search. For example, if the target sex, race, or approximate age of a target is known from a soft biometric label, the set of potential matches to search through can be reduced.

More recent work has explored the capacity to make an identification based on these soft biometrics alone. This has the potential to provide a tremendous advantage in the real world as the soft biometrics can be obtained with no intervention even when the target is at a distance. These, of course, are exactly the conditions in which more traditional biometrics become unavailable due to low resolution or occlusion. Consequently, soft biometrics may offer clear operational value.

The earliest approach using soft biometrics for identification was provided by Samangooei, Guo and Nixon (2008) who asked 38 participants to provide descriptions of ten walkers imaged side-on in the Southampton Gait Database (Shutler, Grant, Nixon & Carter, 2002). The descriptions were selected from a previous study (Macleod, Frowley & Shepherd, 1994) in which participants had an unlimited amount of time to describe a set of target individuals from moving video sequences of the targets walking, or from static photographs taken during the act of walking. A total of 1238 descriptions were extracted, with 1041 providing descriptions of overall physique and the remainder providing descriptions of motion. These were grouped (where possible) and a set of 23 labels was generated covering everyday and easily understandable characteristics such as age, sex, height, etc. Whilst age was represented by 7 categories, and sex by just 2

categories, all other characteristics which could vary continuously, such as height, were reduced to a 5-point scale (i.e., very short, short, average, tall, very tall).. Later, the 13 most reliable labels were incorporated into the final set for (soft biometric) categorical description.

Samangooei et al. (2008) used these 13 labels as prompts to encourage participants to describe ten targets in a process which approximated that of an eyewitness interview. The results suggested substantial agreement across participants when describing two key characteristics – race and sex. Moreover, they revealed significant correlations particularly between those labels that described overall thickness and length of the body, as well as extremities (Samangooei & Nixon, 2014). In particular, common-sense relationships were revealed between body shape and weight, and each correlated with arm thickness, leg thickness and chest descriptions as expected. Consistent with known physiology, a significant correlation was also noted between height and leg length. The lack of a number of other expected correlations may be attributable to the difficulty in describing features when viewed from the side, especially when they related to aspects such as shoulder width which could not easily be discerned.

Of greater interest, however, was the fact that when each of ten targets was compared to stored exemplars in an identification task, Samangooei et al. (2008) revealed a maximum 90% Correct Classification Rate (CCR) based on the soft biometric labels alone. Moreover, this rose to 99.5% CCR when the soft biometric labels were combined with a more traditional biometric method involving automated gait recognition. Consequently, soft biometric labels provided an important input to the identification problem both alone, and when combined with other biometric measures.

1.2 Precision and Comparison

Samangooei et al. (2008) took care to attend to factors that may affect the quality of the soft biometric labels. For example, he allowed participants to view the walkers for as long as needed to ensure that memory constraints did not affect performance. He also asked participants to provide their own values (e.g. when describing height, participants selected one of the five values: Very Short, Short, Average, Tall or Very Tall) rather than constraining their perception by using given labels which may carry connotations that vary across individuals. Finally, he ensured that anchoring issues were minimised by presentation of the walkers in an order that was randomised across participants.

Two factors of concern remained: first, participants' perceptions of others may depend on their expectations of what they consider to be average, and this may vary from one person to the next. For example, what is 'tall' for one person may not be 'tall' for another given their own height. To mitigate against this, Samangooei et al. (2008) obtained participant descriptions of themselves to use as an index reference. However, an inaccuracy of self-report, or an unwillingness to reveal personal information, make this less-than-ideal as a solution.

Second, participants' perceptions of others may suffer through both perceptual and cognitive limitations associated with the perceiver. In particular, the psychological literature has described a phenomenon known as the grain-size strategy (Yaniv & Foster, 1995). This arises because of the dual need for the participant to provide an answer that meets two criteria – accuracy and informativeness (Goldsmith & Koriat, 2008). In an uncertain world in which accuracy cannot be assured, participants use a subjective level of confidence to indicate the likelihood that their answer is correct. When confidence is high, they volunteer the answer and satisfy both accuracy and informativeness. However, when confidence is low, they can still meet a desire to be

accurate by providing an answer that is less fine-grained. For example, they may report a target's age as between 20-40 years rather than between 30-35 years. The concern in the current context is that if a target is sufficiently far away, then estimates of the target characteristics may become so vague that they effectively become useless. Alternatively, participants may simply say 'I don't know' and thus avoid providing a ridiculous answer (Luna, Higham & Martin-Luengo, 2011). Both represent a weakness when soft biometric labels are requested.

One way to address both concerns is to shift away from *categorical* labels (e.g. 'he is tall') towards *comparative* labels ('he is taller than…'). This comparative soft biometric approach was taken by Reid and Nixon (2011) who asked participants to describe one target who was shown alongside another known point of comparison. This procedure avoided the problems associated with individual expectations as participants' judgements were not influenced by their perceptions of themselves but were instead grounded by an objective and known reference. It may also avoid the problems associated with the grain size strategy as comparative judgements may be easier to make than absolute ones. Indeed, the data obtained using categorical labels showed large overlaps between the short, medium and tall labels, suggesting some confusion across participants in the use of the terms. Nevertheless, the categorical labels correlated somewhat with actual walker height as measured from the images (in pixels) (Pearson's correlation = 0.71, $p < .0001$). In contrast, the comparative labels, once sorted, were observed to have a far stronger correlation with walker height (in pixels), suggesting a greater discriminative power from comparative labels than from categorical ones (Pearson's correlation = 0.87, $p < .0001$). Added to this, Reid and Nixon (2013) provided evidence to suggest that the participants themselves far preferred the

comparative method over the categorical method when providing descriptions ($S = 8$, $n = 45$, $p < .01$).

The results using comparative labels are promising. However, the real question of interest is whether identification performance is better when comparative rather than categorical labels are used. In this regard, the data provided by Reid, Nixon and Stevenage (2014) are important. The comparative descriptors were first sorted to derive a rank order of the walkers. From this ranked list, Reid then compared each walker to a database of stored representations simulating an identification task. This revealed a CCR of 95% when comparative descriptors were used, which exceeded the CCR of 90% when categorical descriptors were used previously. Consequently, not only was the comparative approach preferred by the participants, but it yielded labels which were of greater value in the identification task.

The challenge that now presents itself is whether comparative labels may be obtained for the three domains of value in an investigative process – the face, the body and the clothing of a target. The current paper presents data to examine this point. With this in mind, our primary purpose is to obtain descriptions in the form of soft biometric labels for all three domains, in order to determine the relative value of the face, body and clothing on an identification task.

Alongside this piecemeal approach, however, we also evaluate identification performance following the combination of soft biometric labels. This combination, or fusion, of information has only previously been conducted using computer-extracted labels rather than human-generated labels (Arigbabu, Ahmada, Adnan & Yossof, 2015), and results suggested that the combination of face shape, height and body weight improved recognition performance. Surprisingly, however, the addition of computer-extracted labels for skin colour impaired recognition performance. A similar fusion

approach has not yet been conducted using human-generated labels. Accordingly, our second purpose is to evaluate the effect of an intelligent combination, or fusion, of human-generated soft biometric labels with the hope of improving performance on the identification task.

2. Human Description of Body, Face and Clothing

Many approaches have used categorical labels to describe the body and the face (Klare, Klum, Klotnz, Taborsky, Akgul & Jain, 2014; Park, 2010) and have achieved encouraging recognition results on standard databases. In one study using automated facial descriptions (Mery & Bowyer, 2015), the labels[1] were derived by a data-driven approach and were evaluated in the recognition of expressions, gender, race, disguise and beard. A second study (Zhang, Beveridge, Draper, & Phillips, 2015) used estimated gender and race together with face shape. Finally, in the domain of clothing labels, reported results (Jaha & Nixon, 2014) have suggested that they may support identification even when used alone (see also Li, Liu, Wang, Liu & Yan, 2014), and that their combination with traditional soft biometrics allowed a substantial improvement of the otherwise obtained results.

The current work follows from the above work with computer-generated labels or estimations. However, within the current paper, we focus entirely on human-generated labels for the good reason that the human eye is less affected than a camera by factors such as lighting and pose (see Jaha & Nixon, 2015). Additionally, age-related declines in the human are less notable than the degradation that may occur in

[1] The term 'labels' is used throughout the current manuscript. This may be interpreted as being analogous to the term 'attribute' used in the computer vision and biometrics literature cited here.

sources such as over-taped CCTV imagery. Consequently, the current work builds on the recent successful demonstrations with human-generated labels. In particular, it examines the value of such labels (categorical and comparative) when describing the face, body, and clothing of a target.

Given its individuality, it is anticipated that when the face is available for scrutiny, it has the potential to provide a rich vein of information about the target. It is, however, understood that the face may not always be visible to the witness, or to the camera, through disguise (balaclava, motorbike helmet, masks), through occlusion, or through poor resolution. Similarly, when considering the use of clothing as a cue to identity, it is understood that this may provide value over a short time-frame. However, the opportunity to change clothing will affect this as a means of identification over a longer period. Given these assumptions, it is anticipated that the face, when available, will provide the most valuable soft-biometric labels to support identification, followed by the body and then the clothing of the target individual. Moreover, it is anticipated that the combination of soft-biometric labels across face, body and clothing will improve recognition performance beyond the level that is possible when taking each set of labels in isolation.

2.1 Procedure for Generating Labels

2.1.1 Database

The database was comprised of the video sequences depicting 40 targets walking unsupervised along a straight track, in front of a green screen chromakey background. Within each video sequence, the target was visible as a full-length moving figure viewed from the side (Shutler et al., 2002). The chromakey background was used to provide a controlled background which ensured focus entirely on the target. The

majority of the walking targets were young white males (aged around 22 years) and Chinese females (aged around 25 years). The videos were presented in a repetitive loop.

2.1.2 Categorical Body Labelling

A total of 149 participants provided categorical body labels for each of the 40 targets by viewing the video-sequence of each target walking. Video sequences were viewed via a web interface (Samangooei et al., 2008) which allowed the participant to view the video for as long as required. Importantly, each target was viewed one at a time. From this, participants indicated the perceived sex (male, female), and perceived age (Infant, Pre-Adolescence, Adolescence, Young Adult, Adult, Middle Aged or Senior) of the target. In addition, they used the 5-point scales to describe the target for Arm Length, Arm Thickness, Chest, Figure, Height, Hips, Leg Length, Led Direction, Leg Thickness, Muscle Build, Shoulder Shape, and Weight. The categorical soft biometrics labels for the human body were selected for use when a target was at a distance (or at low resolution) and detail could not be perceived. Finally, participants were able to indicate their confidence in their labelling through adjusting the % value associated with their certainty of judgement. However, analyses of the certainty data went beyond the scope of the current study, and is not considered further. Examples of some of the labels used in this study but not their descriptions (terms) are given in Figure 1.

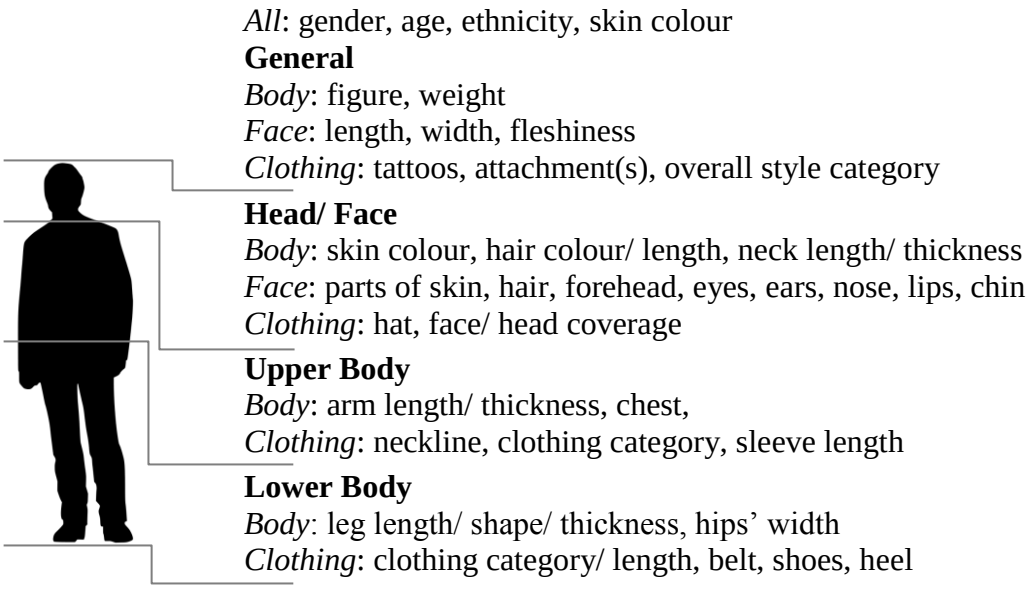


Figure 1. Example labels used for describing the targets

2.1.3 Comparative Face and Body Labelling

A different set of 57 participants provided comparative body and face labels by viewing the 40 targets. For each participant, the target walkers were presented alongside a comparison walker in a web interface designed by Reid et al. (2014) (see Figure 2). As above, the web interface allowed the videos to be replayed for as long as required.

Comparative labels of the body were obtained for the same characteristics as described above. Similarly, comparative labels of the face were obtained for a subset of characteristics outlined in the modified Face Rating Schedule (Sporer, 2007), yielding descriptions of Skin, Hair, Forehead, Eyebrows, Eyes, Ears, Nose, Lips and Chin (for full details, see Reid & Nixon, 2013). Critically, however, all participants gave their descriptions by considering each of the target walkers *relative* to the single comparison walker.

Across all participants, both target and comparison identities were varied so that all 40 target walkers were described, relative to different but known comparisons. This allowed for the final generation of a rank order (A is taller than B is taller than C)

without all comparisons being required. Finally, and as with the categorical labelling procedure, participants were able to indicate their confidence in their labelling through adjusting the % value associated with their certainty of judgement. However, analysis of these data again went beyond the scope of the current study, and is not considered further.

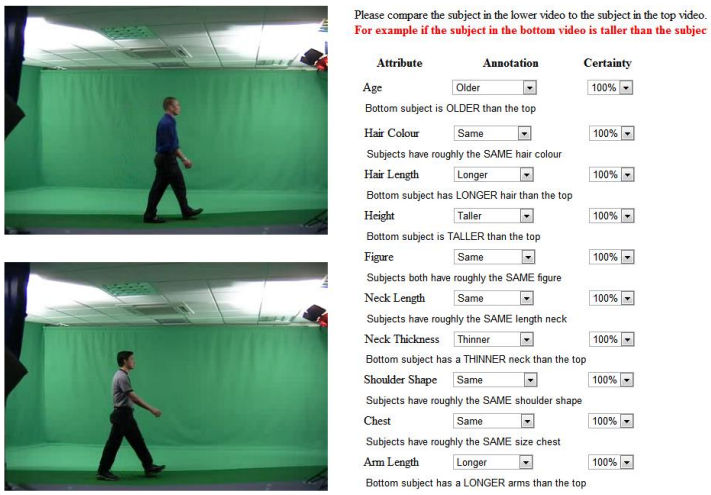


Figure 2: The web interface for deriving comparative body labels (Reid et al., 2014)

2.1.4 Labelling of Clothing

Finally, a different set of 27 participants provided categorical and comparative labels of the clothing of the 40 targets. Participants generated these labels from still photographic images, obtained on the same occasion as the walking video sequences. Thus, clothing and grooming had not altered (see Figure 3). These images were either presented alone (to yield categorical labels), or alongside a known comparison (to yield comparative labels) as above.

The clothing was described using 5-point labels for Head, Upper Body, Lower Body, Foot, Attachments and General Style (for full details, see Jaha & Nixon, 2014). This enabled the generation of a complete description of clothing labels for the whole body, or parts of it. An example of part of the set of categorical and comparative labels to describe clothing for Upper Body, Lower Body and Foot is shown in Figure 4.

(a) clothing

(b) face

Figure 3: Examples images used to generate comparative labels for face and clothing (Jaha & Nixon, 2014; Reid & Nixon, 2013).

Body zone	Semantic Attribute	Categorical Labels	Comparative Labels
Upper body	5. Upper body clothing category	[Jacket, Jumper, T-shirt, Shirt, Blouse, Sweater, Coat, Other]	
	6. Neckline shape	[Strapless, V-shape, Round, Shirt collar, Don't know]	
	7. Neckline size	[Very Small, Small, Medium, Large, Very Large]	[Much Smaller, Smaller, Same, Larger, Much Larger]
	8. Sleeve length	[Very Short, Short, Medium, Long, Very Long]	[Much Shorter, Shorter, Same, Longer, Much Longer]
Lower body	9. Lower body clothing category	[Trouser, Skirt, Dress]	
	10. Shape	[Straight, Skinny, Wide, Tight, Loose]	
	11. Leg length (of lower clothing)	[Very Short, Short, Medium, Long, Very Long]	[Much Shorter, Shorter, Same, Longer, Much Longer]
	12. Belt presence	[Yes, No, Don't know]	
Foot	13. Shoes category	[Heels, Flip flops, Boot, Trainer, Shoe]	
	14. Heel level	[Flat/low, Medium, High, Very high]	[Much Lower, Lower, Same, Higher, Much higher]

Figure 4: Example categorical and comparative labels for clothing (Jaha & Nixon, 2014).

2.2 Overlap between labels

The correlation between facial and body labels, as presented in Figure 5, shows little correlation overall between the two sets of labels (darker cells indicate a lower

correlation) suggesting little overlap. In other words, the processing of gathering descriptive labels for both faces and bodies augments, rather than duplicates, information. This is important because it makes it much more likely that a combination, or fusion, approach will improve recognition performance as additional and non-overlapping information is being added to the mix rather than information that merely repeats already known characteristics.

A similar point is made in the work of O'Toole and colleagues when considering the fusion of face recognition decisions across the quite different approaches taken by human and computer algorithms (Phillips & O'Toole, 2014). However, of more relevance is the work reported by O'Toole, Phillips, Weimer, Roark, Ayyad, Barwick & Dunlop, (2011), who examined performance in an identity matching task when participants were provided with information from either the face, the body, or the two combined. Their results supported our prediction that a combination of inputs would improve performance. Indeed, performance was best when based on the face and body combined. Furthermore, performance was optimised when the inputs were dynamic, as this tended to direct attention to both inputs.

The minimal overlap evident in our correlational matrix bodes well for the fusion analysis to follow. However, this is not to say that there is no overlap whatsoever. In this regard, when comparing descriptions of faces and bodies, it was interesting to note that the strongest correlations appeared between hair colour and descriptors which captured aspects of ethnicity. From Figure 5, hair colour (the Chinese targets invariably had black hair) was highly correlated with skin colour (Skin - Light/Dark). Other labels with a strong correlation to hair colour included Nose (Narrow/Wide), Nose (Flat/Protruding), Eyebrows (Low/High), and Eyes

(Slanted/Round), suggesting that these labels may also be correlated with race and ethnicity.

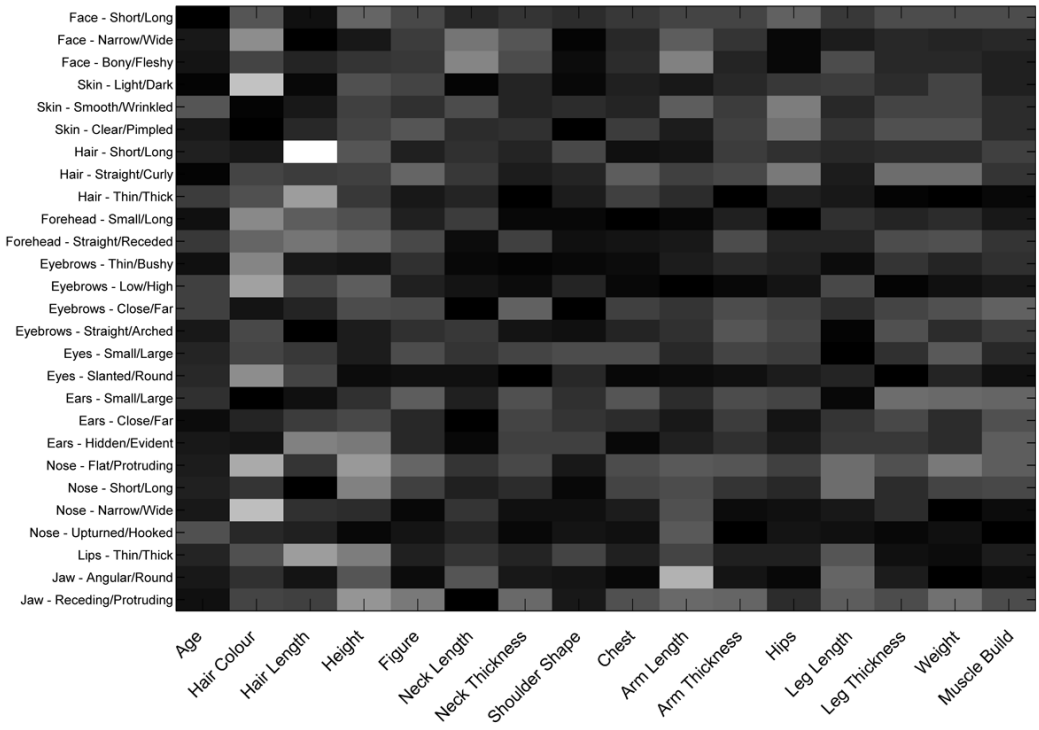


Figure 5. Correlation between facial and bodily comparisons. Lighter cells represent stronger correlations; darker cells represent weaker correlations (Reid & Nixon, 2013).

3. Implementation

3.1 Identification: recognition and verification

The Euclidean distance metric was used to evaluate the similarity between a probe (the target to be identified) and each example within a gallery (the population of known individuals against which the probe is to be compared for identification purposes). This was achieved by ordering the gallery targets, based on their similarity to the probe. In terms of recognition performance, the gallery target with the greatest similarity to the probe was returned as the identity of the probe. However, in terms of verification, the

similarity between the gallery target and the probe had to meet or exceed a pre-selected threshold if it was to be ‘accepted’ rather than ‘rejected’ as a match. Accepting an incorrect target is a False Positive and rejecting a correct target is a False Negative.

A Receiver Operating Curve (ROC) shows the verification rate, which analyses the percentages of False Positives versus False Negatives. A standard and commonly-used measure of performance is the Equal Error Rate (EER), which is the point at which the False Accept Rate (FAR) equals the False Reject Rate (FRR). The lower the EER, the better the performance.

3.2 Feature Vectors and Ranking

For the categorical labels such as sex, the feature vector is formed by a numeric value associated with each term describing that label. The comparative labels need to be sorted into a list which goes from the smallest to the largest (or equivalent) for that label. This list is equivalent to a set of categorical labels from comparative analysis. This list was achieved by using the Elo rating system (Elo, 1978) which was originally aimed to quantify the ranks of chess players. As there is no opportunity for all chess players to play all other players, it is impossible to determine the rank order of players from best to worst through direct comparisons. However, the rank order can be inferred from the results of a partial list of matches against other players. For example, if A beats B, and B beats C, it can be inferred that A would beat C. Similarly, in soft biometrics, the ranks between all targets for each label cannot be observed directly, but may be inferred from a partial list of comparisons.

Taking the chess example, in mathematical terms, a ‘match’ is a comparison between two players, i and j. The match outcome reflects superiority, or not, in performance and hence in status. The outcome is used to adjust the players’ ratings.

Thus, for two players *i* and *j* the ratings R_i and R_j are updated according to the results of a comparison between them. The result of a comparison *S* takes a value 1 for superiority, 0.5 for a tie, and 0 for inferiority (S_i = 0 when player *i* is not superior or equal in comparison), and this is used to update the ratings for two players from iteration <*n*> to iteration <*n*+1> as

$$\begin{aligned} R_i^{<n+1>} &= R_i^{<n>} + k(S_i^{<n>} - E_i) \\ R_j^{<n+1>} &= R_j^{<n>} + k\left(S_j^{<n>} - E_j\right) \end{aligned} \qquad (1)$$

where *E* is the expected outcome given the current ratings. Consequently, the rating is updated by the difference between what has been achieved and what was expected. The parameter *k* is the maximum rating adjustment variable. In the case of soft biometric labels, *k* depends on the available number of comparisons N_C. The maximum rating, *M*, is used to define $k = M/N_C$ allowing *M* to be fully explored by any number of comparisons. *E* is then calculated by

$$\begin{aligned} Q_i &= 10^{R_i/U} \\ Q_j &= 10^{R_j/U} \\ E_i &= Q_i/\left(Q_i + Q_j\right) \\ E_j &= Q_j/\left(Q_i + Q_j\right) \end{aligned} \qquad (2)$$

where *U* is chosen to reflect how a player's current rating affects the expected result. A large value for *U* implies little change to the player's rating, and *U* must exceed zero.

Within the current implementation, the terms describing soft biometric labels were assigned a number in the range -2, 1, 0, +1 or +2 based on their order. The 'score' from a comparison was determined by normalizing the given label's value to within 0 and 1. If the actual result differed little from the expected result then the relative measurements remained unchanged. On the other hand, if the actual result differed considerably from the expected result, the targets' relative measurements were adjusted in the direction indicated by the comparison. The magnitude of adjustment depended on

the difference between the actual and the expected results. In this way, we determined feature vectors for different targets that were comprised of a set of categorical labels together with a set of labels derived by ranking comparative assessments.

2.2.2 Fusing Body, Face and Clothing

In order to analyse recognition performance with respect to possible surveillance scenarios, we investigated recognition performance with different combinations of labels. The Euclidean distance (or match) between two targets *i* and *j* for a feature vector $\boldsymbol{f_i}$ (for target *i*) of *N* measurements $\boldsymbol{f_i} = \{f_{1,i}, f_{2,i} \dots f_{N,i}\}$ is

$$d_{ij} = \sqrt{\sum_{k=1}^{N} \left(f_{k,i} - f_{k,j}\right)^2} \qquad (3)$$

This difference was thresholded, such that a value lower than the threshold represented a match between the targets, whilst a value above the threshold implied no-match. The False Positives and False Negatives were thus derived from the thresholded value. Given that we have three modalities of labels (pertaining to the face, body and clothing of the target), for feature fusion the feature vector becomes a stack of the three modalities. Body *B* with number of features N_B the Body feature vector is $\boldsymbol{f_{i_B}} = \left\{f_{1,i_B}, f_{2,i_B} \dots f_{N_B,i_B}\right\}$ (and Face *F*, Clothing *C*, similarly) and by denoting the modalities as *m* where *m* = 1 for Body, *m* = 2 for Face and *m* = 3 for Clothing, the overall distance by feature fusion is

$$dF_{ij} = \sqrt{\sum_{m=1}^{3} \sum_{k=1}^{N_m} \left(f_{k,i_m} - f_{k,j_m}\right)^2} \qquad (4)$$

Alternatively, we achieve fusion by summation, or by the product rule

$$dP_{ij} = \prod_{m=1}^{3} \sqrt{\sum_{k=1}^{N_m} \left(f_{k,i_m} - f_{k,j_m} \right)^2} \qquad (5)$$

The match scores were then normalised to vary between 0 (match) and 1 (no match), and quality factors were introduced, such as

$$dP\_Q_{ij} = \prod_{m=1}^{3} \sqrt{\frac{\sum_{k=1}^{N_m} \left(f_{k,i_m} - f_{k,j_m} \right)^2}{Q(i,j,m)}} \qquad (6)$$

where the quality was expressed in a probabilistic way as

$$Q(i,j,m,q) = \frac{p\left(d_{i,j_m}, q|C\right)}{p\left(d_{i,j_m}, q|I\right)} \qquad (7)$$

where *C* and *I* were the two possible classes of users, Client (the true subject) and Imposter (a different subject), for a quality factor q. The classification problem was considered using conditional probability for a score **s** by defining a classification as:

$$\begin{gathered} \text{Assign } \mathbf{s} \rightarrow \omega_i, \text{ if} \\ p(\omega_i|\mathbf{s}) > p\left(\omega_j|\mathbf{s}\right), \qquad i \neq j \end{gathered} \qquad (8)$$

where $\boldsymbol{\omega} = \{\omega_1, \omega_2 \ldots \omega_T\}$ and ω_i is the i^{th} class (or target) and T is the number of targets. This formulation of the posterior probability was calculated using the probability density of the score set given a class label given by Bayes theorem as:

$$p(\omega_i|\mathbf{s}) = \frac{p(\mathbf{s}|\omega_i)p(\omega_i)}{p(\mathbf{s})} \qquad (9)$$

where $p(\omega_i)$ is the probability of observing a class, and $p(\mathbf{s})$ is the probability of observing a given score. The class conditional probability $p(\mathbf{s}|\omega_i)$ was the only unknown and was estimated using a parametric technique. The decision is then

$$C_i = \begin{cases} 1 & \prod_m \frac{p(\mathbf{s_m}|\omega_i)p(\omega_i)}{p(\mathbf{s_m})} \leq threshold \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

This approach was applied in two discrete scenarios. The first examined the effect on recognition performance when fusing information across all modalities (the face, body and clothing, where $m = 1,2,3$). In contrast, the second examined the effect on recognition performance when assuming that the face may not be available and thus when fusing information across the remaining modalities (only body and clothing, where $m = 2,3$).

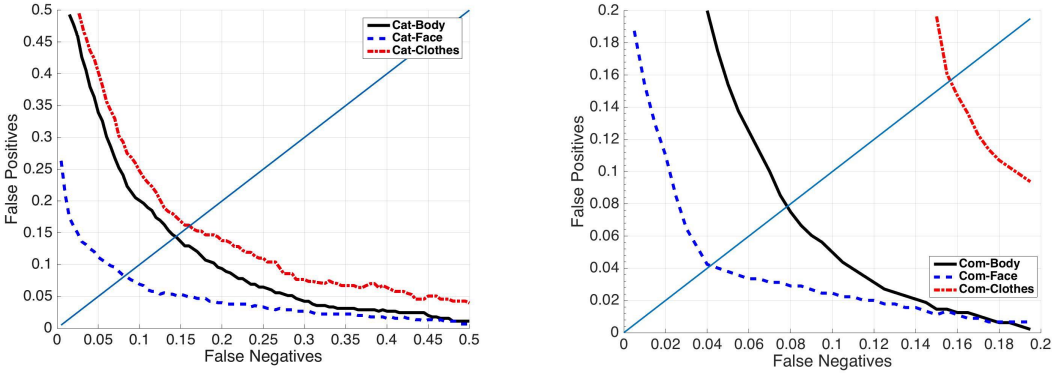
4. Evaluation

4.1 Performance from soft biometric labels for individual modalities

The performance of the three modalities is shown in Figure 6. This revealed several findings of interest. First, the categorical labels (Figure 6a) showed the anticipated order: the face labels offered the best performance (EER = 0.078) followed by the body (EER = 0.136) and lastly the clothing (EER = 0.151). When evaluating the recognition performance associated with the comparative labels (Figure 6b), the overall pattern is similar. However, the performance associated with comparative face labels has improved (EER = 0.052) relative to that based on categorical face labels above. Similarly, the performance associated with comparative body labels (EER = 0.083) has improved relative to that based on categorical body labels. Interestingly, the data suggested fewer False Positives to comparative body labels than to comparative face labels, when the False Negative rate was high.

What was most striking, however, was the very clear result indicating that comparative labels of clothing appeared to be of least utility for recognition with an

EER of 0.155. This said, it should be remembered that clothing labels remain of value when the target is viewed from a distance, as body parts, and especially the face parts, cannot usually be discerned under such conditions. Overall, however, the data here revealed a clear advantage associated with the use of comparative labels over categorical labels and this replicates the results of previous studies (Jaha & Nixon, 2014; Reid &Nixon, 2011).



(a) categorical (b) comparative

Figure 6 Individual performance of three soft biometric modalities

This paper provides the first unified presentation of results by precisely the same process of label generation across the face, body and clothing modalities. Using this superior and well-controlled approach, the results have confirmed the primacy of the face, followed by the body and then the clothing labels when evaluating recognition performance, as demonstrated by Jaha and Nixon (2014), Reid and Nixon (2011) and Samangooei et al. (2008).

4.2 Fusing Soft Biometrics

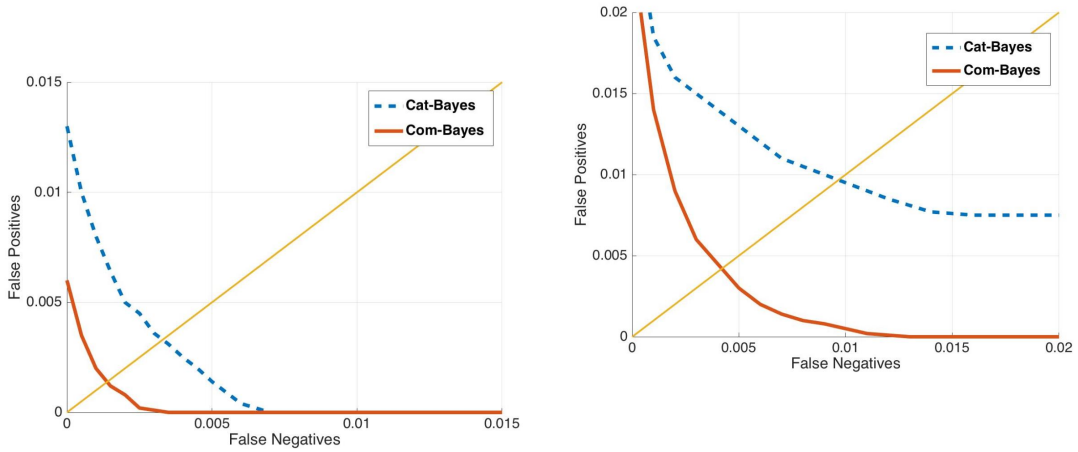
Intuitively, the fusion of face, body and clothing biometrics should improve

performance. This was tested here across two discrete scenarios involving all modalities (face, body and clothing), and involving a plausible subset of modalities (body and clothing only).

When considering the fusion of information across all modalities, the performance in Figure 7(a) showed that the performances based on the fusion of categorical labels was good (EER = 0.0033). However, performance was considerably better when based on the fusion of comparative labels (EER = 0.0014). This latter level of performance was extremely good as the EER is very low and would be acceptable even if it were much larger. Consequently, these data were clear in showing a marked advantage of comparative over categorical labels when fusing information across all three modalities. By comparison to the data reported above, they also showed a marked improvement in performance compared to the levels achieved when based on each modality taken in isolation.

Given the dominance of the face compared to the body and clothing labels reported earlier, one question that remains relates to the level of performance that may be possible if the face becomes unavailable. The second fusion scenario addressed this issue through examining recognition performance when body and clothing information was fused in the absence of face information. This may reflect the real-world situation that exists when viewing a target from such a distance that the face cannot be seen. Figure 7(b) shows the result of fusing body and clothing without including the face. The EER here was 0.0043 and as such, was better than that based on either clothing or body labels when used alone. Indeed, performance was similar to that achieved when using just the face. The dominance of the face implies that, naturally, the higher the resolution the better. However, the current results suggested that if the face could not be seen, it would still be possible to derive identification, with similar accuracy, from a

combination of the body and clothing characteristics. As such, it is not surprising that eyewitness identification forms include face descriptions, but when these cannot be used there remains a rich stock of material for identification (so long as it is exploited).



(a) body, face and clothing (b) body and clothing only

Figure 7: Performance when fusing modalities under two realistic scenarios.

5. Discussion

Society desperately needs ways to identify people from surveillance video: it is common to publish videos or stills of the scene of a crime, and to hope that members of the public can identify the suspects. The current work has explored the potential to make robust identification decisions based on soft-biometric descriptions of the face, body and clothing of an individual. Novelty is provided here through the application of a valuable comparative soft biometric approach in which descriptions are provided relative to an objective standard (comparative labels) rather than relative to some internal norm (categorical labels) which is likely to vary across individuals. In this regard, the results were unequivocal in demonstrating that comparative labels were of greater value than categorical labels in enabling a successful identification of targets. As such, these data support the practice of gathering witness descriptions in a way that

avoids individual differences in expectations, and in a way that minimises vague answers through difficult task demands.

One aspect of using comparative data that has only been alluded to so far is that a number of comparisons is used. Clearly one at least is needed, and fewer are needed for the face than for the body, again reinforcing the superior performance of the face for recognition. Figure 8 shows the influence of the number of comparisons: the face is at near 100% CCR at 5 comparisons, whilst the body reaches 90% CCR at around 10 comparisons. As such the body appears to require at least twice the volume of data needed for the face and still offers lower recognition capability.

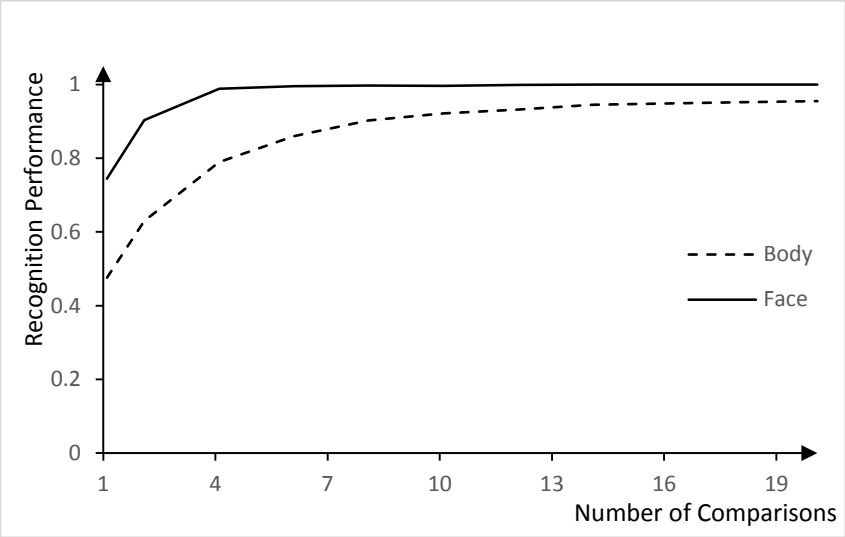


Figure 8 Effect of number of comparisons on recognition performance

5.1 Relative Importance of Face, Body and Clothing

Alongside this demonstration, the data supported the expectation that labels describing the face were of more valuable than those describing the body, which in turn were more valuable than those describing the clothes. As such, the common-sense prediction regarding the primary value of the face as a cue to identity was endorsed here, as was the common-sense prediction of the short-term value of clothing cues. These results sit in contrast to the findings of Lucas and Henneberg (2016) whose recent

work suggested that the body was more valuable than the face when an individual was viewed at a distance. Their method, however, was quite different to that adopted here. They endeavoured to combine cues in a stepwise fashion until there was no-one in their database of 3982 individuals who shared the combination of cues. When such a situation was reached, everyone in the database was said to be individuated. Their results suggested that fewer body cues than face cues were required to reach this point of individuation, hence their conclusion that the body offered more valuable cues for identification than the face.

In accounting for the discrepancy in findings, it is worth noting that Lucas and Henneberg's body and face cues took the form of precise measures (Gordon, Bradtmiller, Churchill, Clauser, McConville, Tebbetts & Walker, 1988) derived by trained individuals given an idealised sample of targets. In contrast, the results reported here were based on descriptions derived from untrained participants. Arguably, the current descriptions may be much more realistic of the data that can rapidly be obtained from a witness about a 'person of interest'. As such, it is possible the body may be more valuable than the face given the time and opportunity to generate precision metrics. However, the face may be more important than the body in the more realistic and ecologically valid conditions where untrained witnesses see unconstrained views of a target.

5.2 The benefit of fusion

Perhaps of greater importance within the present paper is the demonstration of the benefits that are accrued through the intelligent fusion of face, body and clothing information. Indeed, the results of the present paper were clear in demonstrating superior identification performance when descriptors were combined across available sources. Optimal performance was obtained when comparative descriptors of the face,

body and clothing were all combined. This yielded a comparative EER of just .0015 meaning that a false alarm may be expected on only 3 occasions for a population of 2000 targets.

These results mirrored the approach discussed by Yovel and O'Toole (2016) who presented a thoughtful analysis regarding human fusion of dynamic information. They outline a multi-modal model of person perception in which the face, body and voice were combined to provide a rich description of an individual. Indeed, they suggested that identification may usefully be viewed as a process that unfolds over space and time, with the face being the primary cue to identity, but the body and voice being used when the viewing conditions are sub-optimal through distance, pose or occlusion. The results presented here model the benefit shown by humans through integration of different information streams. Interestingly, however, Rice, Phillips Natu, An & O'Toole (2013) suggest that the human onlooker may be unaware of the degree to which they rely on the body or the face when identifying someone. Indeed, participants reported that they relied on facial features to identify someone, and yet eye movement data confirm the use of body cues especially under sub-optimal conditions. The explicit inclusion of each set of descriptors here may have maximised performance through fusion across all information sources available. However, in a real-world witness scenario, human witnesses may need to be prompted regarding soft-biometric descriptors of the body or indeed of clothing in order to unlock their maximal potential.

In this regard, the current paper moves us closer towards an appreciation of person perception as a multidimensional problem and towards identification as a multimodal task. Traditionally, analysis of human capability, and development of automated capacity, has been focussed on one modality at a time – the face, or the voice, or the body, or the distinct style of movement through gait. The current paper,

however, emphasizes the value that results when modalities are combined. Indeed, two important benefits are demonstrated by the current work. First, the fusion approach demonstrates *strength:* when all available information is pooled, identification rates are substantially improved. This showcases optimal performance under ideal conditions when multiple sources of information are available. Second, the fusion approach demonstrates *resilience* against those situations in which only partial information may be available: the combination of partial information still provided superior recognition rates relative to those obtained from each modality taken alone. Indeed, identification was better when body and clothing cues were combined than when taken individually, and this combination yielded performance levels as good as those when the strongest single cue – the face - was considered alone. This result is important as it suggests that even when the face may be unavailable, there remains a rich stock of information contained in the body and the clothing which provides valuable intelligence. This was exactly the scenario provided when trying to identify Jihadi John (who hid his face), and when trying to identify one of the two brothers responsible for the 2013 Boston Marathon bombing (who wore dark glasses and a baseball cap to mask his features) (see Jain & Ross, in press).

6. Considerations of Automated Eyewitness Descriptions

What is clear from the results so far is the fact that powerful automated processing of human descriptions can support a very high level of identification. This raises a complex issue within the legal system regarding the degree to which testimony, whether from witness or expert, may be assisted by an automated system. It is to this issue that the remainder of the discussion is devoted.

The issue of automated assistance assumes importance for several reasons. First, it reflects the discussion provided by Jain and Ross (in press) who emphasized the

return of biometrics research to a forensic setting where it may assist in police investigations and court proceedings. Among the benefits they discuss is the fact that an automated biometric solution can measure values such as confidence intervals, or similarity metrics. These have considerable advantages over the more subjective testimony provided by a witness or an expert in court, not least because they provide a measure of match based on a continuous scale rather than a more blunt, binary, match/no match interpretation (see Champod, 2011; Mnookin, 2008; and Neumann & Champkin, 2012 for useful discussions on the merits of match metrics in the field of fingerprint analysis). Metric-based reporting does, however, depend upon the knowledge of error rates so that judges and jurors can appreciate the value of the information they are provided with. Additionally, it also depends on the non-trivial issue of conveying metric-based information in a way that is accessible to the court. Whilst these are indeed factors that require a response, they should not be factors that prevent a warranted change. Consequently, there is merit in raising the issue of whether the change towards automated assistance, and thus the change towards metrics-based information in court, is warranted.

In this regard, the consideration of automated assistance within investigations and court proceedings is gaining traction given rising concerns over the frailties of the human witness. Human witnesses exhibit problems associated with the processing of an emotional situation including the inability to perceive information as it is, the inability to retrieve information afterwards, and the inability to make reliable use of the information that they have. These human frailties are summarised well by the US National Academy of Sciences Report (2009) which highlights the problem of contextual or confirmatory bias in which humans may perceive information in line with their knowledge, expectations or preconceptions. Human error can arise as a

consequence, and notable examples exist within case law. These include the false arrest of Brandon Mayfield in connection with the Madrid bombing in 2004 (FBI report, 2005), and the fateful shooting of Jean Charles de Menezes who was mistaken as one of the 7/7 bombers involved in the London Underground bombings of 2005. Human errors have also been elicited in the covert testing of experts. Indeed, a set of experts was presented with a pair of fingerprints which they had previously verified as a 'match'. However, they were led to believe they were the Mayfield and Madrid bomber prints and hence 'not a match'. Under these contextually biased circumstances, 4/5 experts changed their original decision to give a wrong answer (Dror, Charlton & Péron, 2006). Together, these results add to the concerns over human error in forensic decision making. Recent research has concentrated on finding ways to minimise such error, through shielding decision makers from irrelevant information, or through encouraging the conduct of blind line-up procedures and blind checking of expert decisions (Haber, 2008). However, the greater involvement of an automated process may also provide a valid way forward.

An interesting discussion of this issue is provided by Dror and Mnookin (2010) who explored the role that an automated system could play within a forensic investigation. They described a novel framework in which man and machine may each contribute to an identification decision but in different ways. Three scenarios may result. First, man and machine may both be capable of the same task in the same way, but the task is offloaded to the machine which thus acts as a 'cognitive servant'. In such an instance, benefits may be felt because the machine may complete the human task in less time, or with greater accuracy given the human tendency to err under a high cognitive load. Second, both man and machine may contribute to the same task in complementary ways in which case the machine is a 'cognitive partner' with each

contributing something that the other cannot contribute. This is the situation described within the current paper, as human raters provided the comparative descriptions, whilst the automated system provided the means for combination and complex analysis to the point of identification. Finally, the machine may offer a superior approach in a task towards which the human has little to contribute. In this case, the machine is the 'cognitive driver'. This latter situation approaches what researchers have referred to as a 'light-out' process, so called because it can effectively be completed even when the lights are turned out and the humans have gone home.

In the current context, a lights-out process would require that the automated system takes over the only part of the task that the humans are currently involved in – the provision of soft biometrics labels describing the perpetrator. If such a capacity could be developed this would mean that witness descriptions could be generated on the basis of CCTV images even if there was no human present to witness the event or to see the perpetrator. This situation is far from being a theoretical notion. Indeed, work is being conducted to explore the accuracy with which soft biometric labels are predicted through automated means (Reid et al., 2014). Success in such an endeavour would enable a fully automated eyewitness statement, based on the intelligent combination of computer-generated identity descriptors – a lights-out solution.

This raises the thorny issue of whether a police investigation, or a court process, would ever accept evidence that has been derived in this way. Such a debate may root itself in ethical issues regarding societal acceptance of a court outcome in which a machine is the bearer of responsibility. Equally, it may raise legal issues regarding the interpretation of expert evidence currently understood as providing either 'fact' or 'opinion'. Whilst the reality may reflect a situation in which automated eyewitness descriptions form part of a case against a defendant but never the entirety of a case, the

ethical and legal issues nevertheless require our scrutiny. What is clear, however, is that we are fast approaching a time when an automated eyewitness statement represents a real possibility, rather than a future prospect.

7. Conclusions and Future Work

It has been known for some time, and with some debate, that eyewitnesses are able to describe people for recognition purposes. Our new study shows that their descriptions can be used within a programme of research aimed towards automated eyewitness descriptions. It is possible to arrange for participants to label targets consistently for recognition using categorical labels of three modalities: the face, the body and the clothing. Of these, the face appears to be the most descriptive modality and leads to the highest recognition capability. Recognition capability is, however, improved when those labels are derived by comparing targets and this improvement is consistent across the three modalities. Moreover, it is possible to fuse these labels for recognition, and this improves performance above that based on any modality when used alone. Current work is aimed to automatically generate the labels by using computer vision techniques, using deep learning and computer vision based methods.

There is a rich field of future research that includes the labels themselves, their generation, their analysis and their uses. There are many extensions that can be made in analysing the performance of the labels (such as by ANOVA, MANOVA – with implicit testing of normality), and it would also be interesting to analyse individual participant performance, especially with respect to factors such as the cross race effect. The labels used for eyes and eyebrows in the face labelling gave some participants cause for concern, and more study could be made of their phrasing and of their potency for recognition. In future, we will need to enlarge the database of targets, to capture a greater variety of appearance in modern societies. For a population of N targets we need

N versions of each categorical label. For the same population we need $\binom{N}{2}$ comparative labels though for full coverage it could be fewer. We continue to investigate ways to use crowdsourcing methods to estimate comparative labels of the body (Martinho–Corbishley, Nixon & Carter, 2016), and of low resolution images of the face (Almudhahkar, Nixon & Hare, 2016), and initial results look promising. Given our focus on fusion within this paper, we will doubtless later use chimeric data, and this is not uncommon in studies on multimodal biometric fusion. Indeed, there is already some exploration of the effect of distance (and image resolution) on the quality of soft biometrics (Tome, Fierrez, Vera-Rodriguez & Nixon, 2014), and this could usefully guide a fusion technique to give greater weight to a modality when distance or resolution factors are optimal. In these ways, we will continue to show how these labels can be used for effective new ways of target recognition, leading to the automatic generation of eyewitness descriptions from images and video.

References

Arigbabu, O.A., Ahmada, S.M.S., Adnan, W.A.W., & Yussof, S. (2015). Integration of multiple soft biometrics for human identification, *Pattern Recognition Letters*, 68(2), 278–287.

Almudhahka, N., Nixon, M.S., & Hare, J.S. (2016). Human face identification via comparative soft biometrics, *Proc. IEEE ISBA.*

Champod, C. (2011). Fingerprint examination: Towards more transparency. *Law, Probability and Risk. 7,* 165-189.

Dantcheva, A., Elia, P., & Ross, A. (2016). What else does your biometric data reveal? A survey on soft biometrics, *IEEE Trans. on IFS*, 11(3), 441-467.

Dror, I.E., Charlton, D., & Péron, A.E. (2006). Contextual information renders experts vulnerable to making erroneous identifications, *Forensic science international*, 156(1), 74–78.

Dror, I.E., & Mnookin, J.L. (2010). The use of technology in human expert domains: challenges and risks arising from the use of automated fingerprint identification systems in forensic science, *Law, Probability & Risk*, 9(1), 47-67.

Elo, A.E., (1978). *The rating of chessplayers, past and present*, Batsford.

Federal Bureau of Investigation (1985). The science of fingerprints: Classification and uses. Washington, DC: US Government Printing Office.

Goldsmith, G., & Koriat, A. (2008). The strategic regulation of memory accuracy and informativeness. In A Benjamin & B Ross (Eds.). *The Psychology of Learning and Motivation*, Vol 48: Memory use as skilled cognition (pp 307-324). San Diego, CA: Elsevier.

Gordon, C.C., Bradtmiller, B., Churchill, T., Clauser, C.E., McConville, J.T., Tebbetts, I.O., & Walker R.A. (1988). *Anthropometric Survey of US Army Personnel: Methods and Summary Statistics*, Technical Report NATICK/TR-89/044, United States Army

Haber, R.N., & Haber, L. (2008). Scientific validation of fingerprint evidence under Daubert. *Law, Probability & Risk*, 7, 87-109.

Jaha, E.S., & Nixon M.S. (2014). Soft biometrics for subject identification using clothing attributes. *Proc. Int. Joint Conf. on Biometrics, IJCB.*

Jaha, E.S., & Nixon M.S. (2015). Viewpoint Invariant Subject Retrieval via Soft Clothing Biometrics. *Proc. Int. Conf. on Biometrics, ICB.*

Jain, A.K., & Ross, A. (in press). Bridging the gap: from biometrics to forensics. Philosophical Trans. Roy. Soc. B, DOI: 10.1098/rstb.2014.0254

Klare, B., Klum, S., Klontz, J., Taborsky, E., Akgul, A., & Jain, A.K. (2014). Suspect Identification Based on Descriptive Facial Attributes. *Proc. Int. Joint Conf. on Biometrics, IJCB.*

Li, A., Liu, L., Wang, K., Liu, S., & Yan, S. (2014). Clothing Attributes Assisted Person Re-identification, *IEEE Trans. on Circuits and Systems for Video Technology*, 25(5), 869-878, 2015.

Lucas, T., & Henneberg, M. (2016). Comparing the face to the body, which is better for identification? *Int. J. Legal Medicine*, 130(2), 533-540.

Luna, K., Higham, P.A., & Martin-Luengo, B. (2011). Regulation of memory accuracy with multiple answers: The plurality option. *Journal of Experimental Psychology: Applied*, 17(2), 148-158.

MacLeod, M.D., Frowley, J.N., & Shepherd, J.W. (1994). Whole body information: Its relevance to eyewitnesses, Adult Eyewitness Testimony, 6, Cambridge University Press.

Martinho-Corbishley, D., Nixon, M.S., & Carter, J.N. (2016). Analysing Comparative Soft Biometrics from Crowdsourced Annotations, *IET Biometrics*, 5(4), 276–283.

Mery, D., & Bowyer, K.W. (2015). Automatic facial attribute analysis via adaptive sparse representation of random patches, *Pattern Recognition Letters*, 68(2), 260–269.

Mnookin, J.L. (2008). The validity of latent fingerprint identification: Confessions of a fingerprinting moderate. *Law Probability and Risk, 7*, 127-141.

National Academy of Sciences (2009). *Strengthening forensic science in the United States: A path forward*. Washington, DC. National Academies Press.

Neumann, C., & Champkin, J. (2012). Fingerprints at the crime-scene: Statistically certain or probably? *Significance*: *The Royal Statistical Society, February 2012,* 21-25.

Nixon, M.S., Correia, P.L., Nasrollahi, K., Moeslund, T.B., Hadid, A., & Tistarelli, M. (2015). On soft biometrics, *Pattern Recognition Letters*, 68(2), 218–230.

O'Toole, A.J., Phillips, P.J., Weimer, S., Roark, D.A., Ayyad,J., Barwick, R., & Dunlop, J. (2011). Recognizing people from dynamic and static faces and bodies: Dissecting identity with a fusion approach. *Vision Research,* 51(1), 74-83. doi: 10.1016/j.visres.2010.09.035

Park, U., & Jain, A.K. (2010). Face Matching and Retrieval Using Soft Biometrics, *IEEE Trans on Information Forensics and Security*, 5(3), 406-415.

Phillips, P.J., & O'Toole, A.J. (2014). Comparison of human and computer performance across face recognition experiments. *Image and Vision Computing, 32(1),* 74-85. doi: 10.1015/j.imavix.2013.12.002

Reid, D.A., & Nixon, M.S. (2011). Using Comparative Human Descriptions for Soft Biometrics, *Proc. Int. Joint Conf. on Biometrics, IJCB.*

Reid, D.A., & Nixon, M.S. (2013). Human identification using facial comparative descriptions, *Proc. Int. Conf. on Biometrics, ICB.*

Reid, D.A., Nixon, M.S., & Stevenage, S.V. (2014). Soft biometrics; human identification using comparative descriptions. *IEEE Trans. on PAMI*, 36(6), 1216-1228.

Rice, A., Phillips, P.J., Natu, V., An, X., & O'Toole, A.J. (2013). Unaware person recognition from the body when face identification fails. *Psychological Science, 24*, 2235-2243

Samangooei, S., Guo, B., & Nixon, M.S. (2008). The use of semantic human description as a soft biometric. *Proc. Int. Conf. on Biometrics: Theory, Applications and Systems, BTAS.*

Samangooei, S., & Nixon, M.S. (2014). On Semantic Soft-Biometric Labels, *Proc. BIOMET 2014.*

Shutler, J.D., Grant, M.G., Nixon, M.S., & Carter, J.N. (2002). On a large sequence-based human gait database, *Proc. RASC.*

Sporer, S.L. (2007) Person Descriptions as Retrieval Cues: Do they really help?, Psychology, Crime & Law, 13(6): 591-609

Tome, P., Fierrez, J., Vera-Rodriguez, R., & Nixon, M.S. (2014), Soft Biometrics and their Application in Person Recognition at a Distance. *IEEE Trans. on IFS*, **9**(3):464-475,

Yaniv, I., & Foster, D. (1995) Graininess of judgement under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General*, 124, 424-432.

Yovel, G., & O'Toole, A.J. (2016). Recognizing People in Motion, Trends in Cognitive Sciences, 20(5), 383–395.

Zhang, H., Beveridge, J.R., Draper, B.A., & Phillips, P.J. (2015). On the effectiveness of soft biometrics for increasing face verification rates, *CVIU*, 137:50–62.

Figure 2 enlarged

Figure 6(a) enlarged

Body zone	Semantic Attribute	Categorical Labels	Comparative Labels
Upper body	5. Upper body clothing category	[Jacket, Jumper, T-shirt, Shirt, Blouse, Sweater, Coat, Other]	
	6. Neckline shape	[Strapless, V-shape, Round, Shirt collar, Don't know]	
	7. Neckline size	[Very Small, Small, Medium, Large, Very Large]	[Much Smaller, Smaller, Same, Larger, Much Larger]
	8. Sleeve length	[Very Short, Short, Medium, Long, Very Long]	[Much Shorter, Shorter, Same, Longer, Much Longer]
Lower body	9. Lower body clothing category	[Trouser, Skirt, Dress]	
	10. Shape	[Straight, Skinny, Wide, Tight, Loose]	
	11. Leg length (of lower clothing)	[Very Short, Short, Medium, Long, Very Long]	[Much Shorter, Shorter, Same, Longer, Much Longer]
	12. Belt presence	[Yes, No, Don't know]	
Foot	13. Shoes category	[Heels, Flip flops, Boot, Trainer, Shoe]	
	14. Heel level	[Flat/low, Medium, High, Very high]	[Much Lower, Lower, Same, Higher, Much higher]

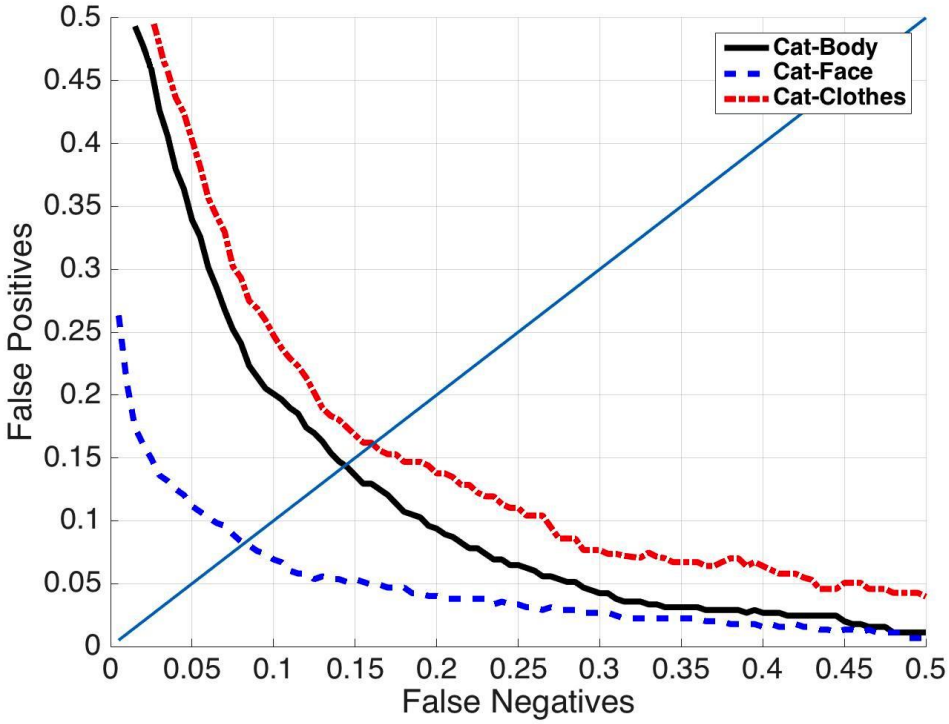


Figure 6(b) enlarged

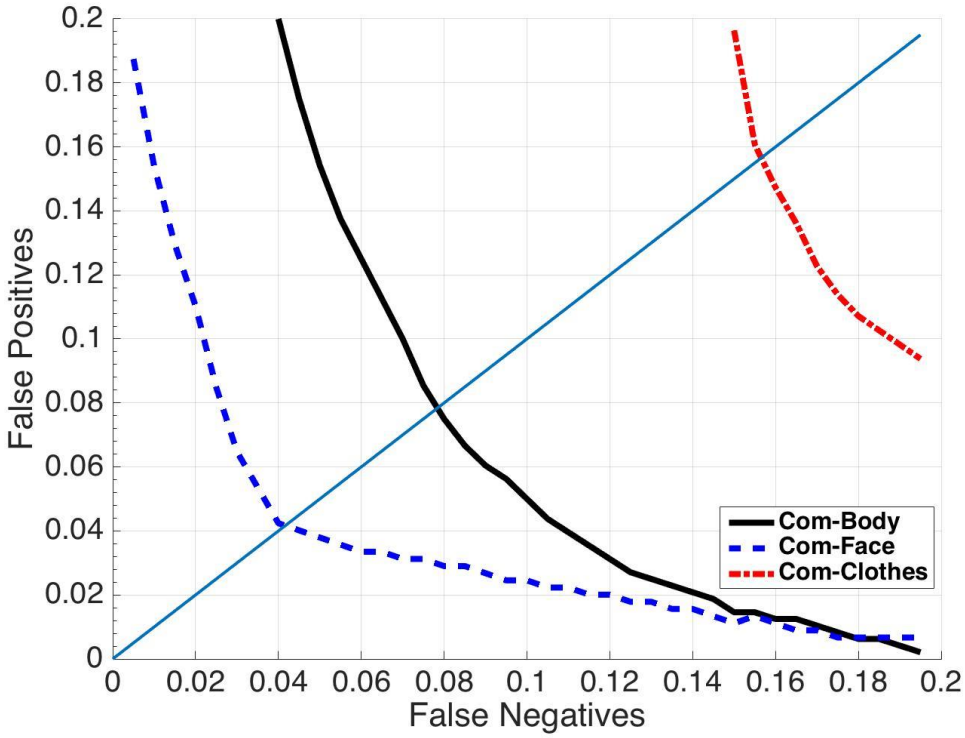


Figure 7(a) enlarged

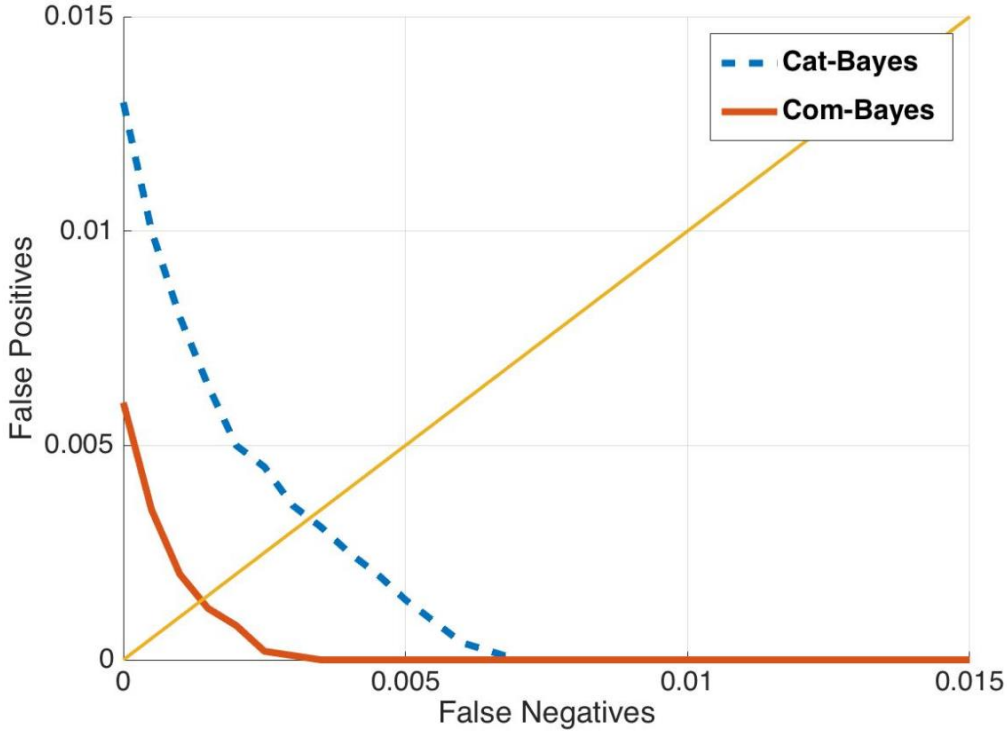


Figure 7(b) enlarged