

# Soft Biometric Retrieval to Describe and Identify Surveillance Images

Daniel Martinho-Corbishley, Mark S. Nixon and John N. Carter,  
School of Electronics and Computer Science,  
University of Southampton, United Kingdom.  
{dmc,msn,jnc}@ecs.soton.ac.uk

## Abstract

*Soft biometrics are human describable, distinguishing human characteristics. We present a baseline solution to the problem of identifying individuals solely from human descriptions, by automatically retrieving soft biometric labels from images. Probe images are then identified from a gallery of known soft biometric signatures, using their predicted labels. We investigate four labelling techniques and a number of challenging re-identification scenarios with this method. We also present a novel dataset, SoBiR, consisting of 8 camera viewpoints, 100 subjects and 4 forms of comprehensive human annotation to facilitate soft biometric retrieval. We report the increased retrieval accuracy of binary labels, the generalising capability of continuous measurements and the overall performance improvement of comparative annotations over categorical annotations.*

## 1. Introduction

Biometric verification is now prevalent in today's society, through facial recognition passport programmes and mobile devices supporting fingerprint recognition. While humans are able to effortlessly *describe* and *identify* each other through visual cues, performing *automatic visual identification* still proves challenging. We seek to identify individuals from video surveillance footage using only human descriptions obtained from eye witness testimonies, surveillance operators and 'super-recognisers'.

**Problem.** The topic of re-identification has gained much recent interest, yet a truly generalised solution is still to be proposed. Almost all approaches that claim high recognition rates do so by matching images directly to one another [14, 2], or by fusing auxiliary semantic attributes with low-level image features [29, 1, 13]. Though these provide state-of-the-art re-identification performance on prevalent benchmark datasets, their scope is limited to situations where an image of the probe subject is available from a directly corresponding camera and environment.



Figure 1: Pre-processed SoBiR image viewpoint examples.

However, surveillance operators often need to search video footage given only a *human description* of a suspect. Therefore, we need a novel solution to automatically describe and identify suspects from images, without relying on image matching techniques or hard biometrics.

Our problem encapsulates three main challenges: dealing with the enormous variability in CCTV images, improving accessibility for human operators and investigating a representative variety of surveillance scenarios.

**Proposal.** To move towards solving these problems, we advocate performing recognition solely in a 'semantic space' that is understood by human operators, in order to provide better generalised and widely applicable solutions.

We propose to automatically retrieve global and body *soft biometric traits* from images, which are then used to match and identify subjects. Unlike traditional hard biometrics, soft biometrics are formed as a collection of human describable, distinguishing visual features [19, 5, 4]. *Soft traits* (e.g gender, height etc.) are derived from salient and permanent, global and body characteristics [19, 25, 24, 16], setting them apart from other mid-level semantic attributes.

Soft biometrics can also be deployed where other biometrics cannot, having stronger invariance to human pose, camera distance and illumination changes [19, 11].

We present a novel dataset with which to investigate several re-identification scenarios, including both images [26] (Figure 1) and a comprehensive set of soft biometric annotations [25, 16]. Other widely used benchmark datasets (e.g. VIPeR [9] and GRID [15]) only allow ‘one-shot’ re-identification results to be reported from one pair of cameras. By exploring a number of varied re-identification scenarios, we can better understand the generalising capability and discriminative power of our techniques.

Specifically, we investigate absolute binary, absolute categorical, relative binary and relative continuous labelling techniques, derived from two sources of soft biometric annotation (Table 1). We can therefore compare the strengths and weaknesses of each labelling and annotation method when performing retrieval and recognition.

Our contributions are as follows:

**The publicly accessible SoBiR dataset**<sup>1</sup>, cataloging 100 subjects captured from 8 camera viewpoints, comprising of 1600 images, described using 12 soft biometric traits through categorical annotations, comparative annotations and derived binary and continuous labels.

**A baseline approach** to soft biometric retrieval and recognition for four labelling methods, including soft trait optimisation.

**A performance analysis** of each labelling method, in traditional one-shot, multi-shot, disjoint re-identification and zero-shot identification scenarios, providing evidence of the improved performance using comparative annotations over categorical annotations.

## 2. Related work

Our work follows several fields, most notably soft biometric recognition and pedestrian re-identification, alongside image attribute classification and pairwise ranking techniques.

**Soft biometric recognition.** Nixon et al. and Dantcheva et al. provide comprehensive surveys exploring the topic of soft biometrics [19, 4]. Earlier soft biometric studies focus on annotating a set of absolute categorical labels, analysing the mutual exclusivity and identifying ability of each soft trait [25, 5]. Later studies collect comparative annotations, by comparing pairs of subjects in a more objective manner [24, 16, 11]. Our dataset and investigation incorporates both forms of annotation. Dantcheva et al. is one of the earliest works to perform soft biometric retrieval and identification, showing promise for future work [5].

Annotation	Measure	Label type	Combinations	Balanced
Categorical	Absolute	Multi-class	23592960	No
Categorical	Absolute	Binary	4096	No
Comparative	Relative	Continuous	$\infty$	-
Comparative	Relative	Binary	4096	Yes

Table 1: Labelling method characteristics.

**Relative attributes.** In order to classify image attributes relative to one another, Parikh et al. proposes an extension to the RankSVM algorithm including similarity constraints [20]. As a consequence, many studies use this algorithm to perform subject ranking given pairwise comparisons [17, 11], and it is used to model our dataset’s relative continuous and relative binary labels, following Martinho et al. [16].

**Attribute classification.** There have been many works on classifying attributes directly from images, for facial verification [12], demography from face [10], gender estimation from face and body [18], describing faces and scenes [20], describing clothing [11] and describing texture [17]. A large portion of related research is also dedicated to recognising attributes from pedestrians and surveillance footage [29, 1, 21, 13].

**Attribute re-identification.** We extend Layne et al.’s ‘zero-shot’ identification scenario, where only a probe semantic description is available for identification, without an image [13]. We also implement similar low-level feature descriptors to Gray et al. and Prosser et al. [9, 23].

Recent advances in neural networks have lead to even more improved human attribute classification approaches [21, 29]. Zhu et al. propose a multi-label convolution neural network to jointly predict 21 attributes, also reporting re-identification results fused with image features [29].

**Re-identification challenges.** Gong et al. discuss the challenges and limitations of ‘closed-world’ datasets as opposed to ‘open-world’ re-identification scenarios [8]. We explore some of the questions raised in this paper, with regards to inter- and intra-class variation, small numbers of samples per subject, and generalising capability across cameras, using our relatively small, but highly detailed dataset of 100 subjects. Vezzani et al. highlight the application benefits of measuring a biometric profile, enabling both re-identification and biometric recognition [27]. Finally, Wang et al. present a strong argument for using semantic features, stating qualities such as: viewpoint independence, robustness and capacity for human interpretation [28].

## 3. Soft Biometric Retrieval Dataset

The Soft Biometric Retrieval (SoBiR) dataset is designed to be a challenging, varied and flexible framework with

<sup>1</sup>SoBiR1.0 is available at <http://users.ecs.soton.ac.uk/dmclg14/#isba-16>

Trait	Comparative [16] [More A, More B]	Categorical [25] [0, 1, 2, 3, 4, 5, 6, ...]
Gender	[Feminine, Masculine]	[Female, Male]
Age	[Old, Young]	[Infant, Pre Adolescence, Adolescence, Young Adult, Adult, Middle Aged, Senior]
Height	[Tall, Short]	[Very Short, Short, Average, Tall, Very Tall]
Weight	[Heavy, Light]	[Very Thin, Thin, Average, Big, Very Big]
Figure	[Fat, Thin]	[Very Small, Small, Average, Large, Very Large]
Chest size	[Big, Small]	[Very Slim, Slim, Average, Large, Very Large]
Arm thickness	[Thick, Thin]	[Very Thin, Thin, Average, Thick, Very Thick]
Leg thickness	[Thick, Thin]	[Very Thin, Thin, Average, Thick, Very Thick]
Skin colour	[Dark, Light]	[White, Tanned, Oriental, Black]
Hair colour	[Dark, Light]	[Black, Brown, Red, Blond, Grey, Dyed]
Hair length	[Long, Short]	[Shaven, Short, Medium, Long]
Muscle build	[Muscle, Lean]	[Very Lean, Lean, Average, Muscly, Very Muscly]

Table 2: Lexicon of comparative and categorical traits and labels included in SoBiR.

which to perform soft biometric retrieval for person re-identification. Rather than pursuing ever larger image collections, datasets like SoBiR are essential for the exploration of diverse and innovative re-identification solutions and scenarios. Using SoBiR’s highly detailed soft biometric descriptions opens up new avenues for re-identification solutions that are not constrained to image matching.

**Image set.** The dataset images are derived from a subset of subjects and images captured as part of the Southampton University Multi-Biometric Tunnel dataset [26]. The original dataset captured images of subjects walking along a multi-biometric tunnel, from 6 pairs of cameras. We select 2 images for all 100 subjects, sampled at random distance from 4 pairs of cameras; *front*, *back*, *side* and *top* views, exhibiting variations in pose and image quality (Figure 1). Camera placements aim to replicate a high number of surveillance viewpoints as subjects move through the tunnel, capturing both left and right-hand sides in a constrained environment.

Each camera pair can be used to perform traditional ‘one-shot’ re-identification, similarly to VIPeR and GRID. Our images have more consistent lighting and higher resolution, pre-conditioned with basic background subtraction.

An *all* view set combines image samples from each camera, enabling multi-shot re-identification, disjoint re-identification and zero-shot identification scenarios. We supply two dataset forms, one with only background subtraction applied, original image dimensions and no other-preprocessing. The second is a pre-processed image set, used for the retrieval experiments in this report.

**Image pre-processing.** Maintaining their aspect ratio, images are first scaled to a height of 256 pixels. A horizontal mid-point is calculated for each image, taking the median value of each row’s mean non-white pixel location. Scaled images are then placed in a 256x256 white square, aligning the horizontal mid-point to 128 pixels across. Any image pixels falling outside the 256x256 area are cropped. Lastly, adaptive histogram equalisation is applied [22].

**Soft biometric annotations.** We draw the ground truth soft biometric labels from two sources of human annotation. Table 2 details the 12 soft traits and corresponding lexicons originally annotated for our dataset’s 100 subjects. Table 1 outlines the four modes of labelling analysed in this report.

*Absolute categorical* annotations are presented by Samangoee et al., collecting a set of 23 global, body and head traits using a multi-class annotation system [25]. *Absolute binary* labels are derived from the categorical multi-class annotations by combining each label class into two semantic groups, e.g. shorter and taller, lighter and darker etc. Binary groupings are formed such that the number of annotations for the new binary labels are as equally balanced as possible.

The second set of annotations are provided by Martinho et al., collecting pairwise subject comparisons, through an online crowdsourcing task using a 5-point, bi-polar scale [16]. *Relative continuous* values are derived from the pairwise annotations using the RankSVM formulation with similarity constraints, detailed by Parikh et al. [20]. Normalised values (between 0.0 and 1.0) are then split at 0.5, forming perfectly balanced *relative binary* labels. Though these new binary annotations are more coarse, they are still *relative measures* in their own right.

## 4. Approach

In this section we present a baseline approach for automatic soft biometric label retrieval and recognition. First, we clarify some key definitions:

**Retrieval** is the process of learning and predicting soft biometric labels or signatures from *probe* images.

**Recognition** is the process of matching soft biometric signatures, between previously unknown probe subjects and known *gallery* subjects.

**Identification** is the amalgamation of retrieval and recognition, to generate probe signatures, then identify them from a gallery of known subject signatures.

We retrieve soft biometric signatures from images by training an ensemble of prediction algorithms. Soft biometric labels are learnt between ground-truth annotations

and low-level image features, generated by hand-crafted descriptors. During the first round of training we search for predictor parameters and soft trait weightings to improve retrieval and optimise recognition stages respectively.

#### 4.1. Image feature descriptor

We generate a 4704 length feature vector, from each pre-processed 256x256 image, taking inspiration from the Ensemble of Localised Features (ELF) descriptor [9], and Prosser et al.’s horizontal strips [23].

Images are split into six equally sized horizontal strips, each described by 9 colour channels; RGB, HSV, Lab (CIELAB), and 16 Gabor [6] luminance channel texture filters;  $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$ ,  $\sigma \in \{1, 3\}$ ,  $\lambda \in \{0.05, 0.25\}$ . Channels are represented using 16 bins, producing a total of 400 features per strip. A Histogram of Oriented Gradients descriptor [3] is applied to the entire image, using 9 orientations, 16x16 pixels per cell and 1 cell per block, represented in a further 2304 features.

Images are mirrored vertically at run-time, generating a second descriptor to provide higher viewpoint invariance.

#### 4.2. Soft Biometric Retrieval

We train  $M = 1, \dots, 12$  learning functions to predict soft biometric labels  $\mathbf{y} = \{y_i^m\}$ , given a set of training image features  $\{\mathbf{x}_i\}$  and ground-truth annotations  $\mathbf{a} = \{a_i^m\}$ , where  $m \in M$ . We use the Extra-Trees (ET) supervised ensemble learning algorithm [7], for its reduced variance and computational efficiency. For binary and multi-class (categorical) labelling we use an ET classifier, while for continuous labelling we use an ET regressor.

During the first round of training we perform a parameter grid search with 2-fold cross validation. The best parameter set for each soft trait predictor is chosen by evaluating the loss function norm  $\|L\|$ . We search for both the number of estimators  $n_{est} \in \{1, 5, 10, 50, 100, 200, 300, 400\}$  and maximum features  $n_{max} \in \{auto, sqrt, log\}$ . For binary and multi-class labelling, we define  $L$  as the Hamming loss function between ground-truth and predicted labels;

$$L(\mathbf{y}, \mathbf{a}) = \left( \frac{1}{n^m |I|} \sum_{i \in I} xor(y_i^m, a_i^m) : m \in M \right), \quad (1)$$

where  $n^m$  is the number of classes for soft trait  $m$ . For continuous labelling, we define  $L$  as the mean squared error;

$$L(\mathbf{y}, \mathbf{a}) = \left( \frac{1}{|I|} \sum_{i \in I} (y_i^m - a_i^m)^2 : m \in M \right). \quad (2)$$

#### 4.3. Soft Biometric Recognition

Given  $\{\mathbf{z}_p\}$  retrieved soft biometric signatures, we aim to recognise all probe subjects  $p \in P$ , from the gallery of  $G = 1, \dots, 100$  subject signatures  $\{\mathbf{s}_g\}$ , where  $g \in G$ .

Experiment	No. Samples		No. Subjects		No. Cameras	
	Tr.	Te.	Tr.	Te.	Tr.	Te.
One-shot re-id.	100	100	100	100	1	1
Basic re-id.	700	100	100	100	7	1
Disjoint re-id.	600	100	100	100	6	1
Zero-shot id.	728	144	91	9	8	8

Table 3: Experiment train (Tr.) and test (Te.) set statistics.

Soft traits do not have equal discriminative ability (discussed in Section 5.1), being affected by retrieval prediction accuracy, ground-truth annotation method and measurement precision. Therefore, in order to optimise recognition, we weight each soft trait. We introduce an objective rank minimisation function to find weight vector  $\mathbf{w}$ , where  $|\mathbf{w}| = M$ , run on the first round of training;

$$O = \sum_{p \in P} \left( \sum_{g \in G} \begin{cases} 1, & \text{if } \|\mathbf{w}^T L(\mathbf{z}_p, \mathbf{s}_g)\| < \|\mathbf{w}^T L_p\| \\ 0, & \text{otherwise} \end{cases} \right)^\lambda \quad (3)$$

where  $L_p$  is the loss vector between the predicted and ground-truth signatures for probe  $p$ . By setting  $0 < \lambda < 1$ , we give a slight precedence to reducing already low ranks over higher ranks. We empirically choose  $\lambda = 0.8$ . Rather than attempting to improve the average rank, or simply rank 1 for all samples, our formulation produces robust solutions, even when allowed only one training iteration. This method is an alternative to Layne et al.’s expected rank formulation [13]. To prevent overfitting, we initialise  $\mathbf{w}$  to randomly distributed values between 0.5 and 1.5.

### 5. Experiments

To showcase our baseline soft biometric identification methods and the extent of the SoBiR dataset, we explain and analyse two sets of experiments. The goal of each experiment is to assess how accurately our retrieval methods can identify probe suspects. A probe soft biometric signature is predicted from an unseen image and matched against the gallery of all known soft biometric subject signatures.

All experiments follow the same format after defining a train-test split criteria (Table 3). Train-test sets are formed through 11-fold cross validation. The first fold is used as a setup stage for learning soft trait predictor parameters and optimisation weights. Results are reported as an average of the remaining 10 folds, by matching retrieved soft biometric signatures against ground-truth annotations.

**One-shot re-identification.** We perform one-shot re-identification across the initial four pairs of cameras, observing the performance of our four labelling methods in a more traditional environment. In each experiment, the dataset consists of one camera pair, resulting in two images



Labelling	r=1	r=5	r=10	r=25	r=50	r=75	nAUC	r+
<b>(a) One-shot re-identification (average across camera pairs)</b>								
Abs. Cat.	9.3	29.4	43.8	68.5	88.8	96.7	78.9	5.1
Abs. Bin.	8.4	27.9	43.0	69.9	89.6	97.2	79.4	4.9
Rel. Bin.	<b>12.5</b>	<b>36.0</b>	<b>53.7</b>	79.3	94.0	99.0	84.4	1.5
Rel. Cont.	9.4	34.9	53.5	<b>80.4</b>	<b>95.4</b>	<b>99.3</b>	<b>84.8</b>	4.9
<b>(b) Multi-shot re-identification</b>								
Abs. Cat.	15.7	40.8	54.3	77.1	92.5	98.0	83.4	7.1
Abs. Bin.	13.3	37.0	53.9	75.2	93.1	97.9	83.2	6.1
Rel. Bin.	<b>20.1</b>	<b>49.5</b>	<b>77.1</b>	<b>86.3</b>	<b>95.4</b>	<b>99.4</b>	<b>88.1</b>	1.0
Rel. Cont.	10.3	32.4	50.5	79.2	94.8	99.1	84.2	3.8
<b>(c) Disjoint re-identification</b>								
Abs. Cat.	12.2	32.2	45.0	67.9	85.5	95.9	78.1	6.0
Abs. Bin.	8.3	26.9	43.3	67.5	86.6	96.5	78.2	5.4
Rel. Bin.	<b>13.1</b>	<b>37.0</b>	<b>55.2</b>	<b>77.5</b>	<b>91.6</b>	<b>98.0</b>	<b>83.1</b>	0.9
Rel. Cont.	7.5	31.1	47.1	74.4	91.5	97.6	81.5	3.7
<b>(d) Zero-shot identification</b>								
Abs. Cat.	1.4	8.6	18.1	36.3	64.5	87.8	60.4	3.9
Abs. Bin.	1.7	15.2	<b>28.2</b>	48.4	71.1	90.0	66.5	3.5
Rel. Bin.	2.6	15.8	26.4	52.3	<b>81.4</b>	<b>94.1</b>	<b>70.8</b>	2.3
Rel. Cont.	<b>6.7</b>	<b>16.4</b>	27.9	<b>53.9</b>	77.4	93.6	70.2	4.4

Table 4: Recognition scenario CMCs % (top values emboldened), **nAUC** denotes normalised Area Under Curve (%), **r+** denotes average rank optimisation.

per subject. Train-test sets each contain one image per subject, randomly sampled from alternate cameras.

**Assorted re-identification scenarios.** The second set of experiments exemplify simulations of more real-world scenarios with SoBiR. We now use all 8 cameras from the *all* view set, resulting in 8 images per subject. *Multi-shot re-identification* randomly samples 1 image per subject for the test set, allocating the remaining 7 to the training set. *Disjoint re-identification* randomly samples 1 image per subject, but allocates only 6 of the remaining images to the training set, excluding the test set’s alternate camera from the training set. In this way, probe images will vary highly to the training images for a corresponding subject. Finally, *zero-shot identification* is the most challenging scenario, simulating an eye witness description of a subject, without relevant training image examples. Train-test sets are split across subjects, with all images for 9 randomly sampled subjects allocated to the test set and remaining 91 subjects’ images allocated to the training set.

### 5.1. One-shot re-identification results

Table 4a and Figure 2a reveal how relative labelling methods outperform absolute methods during one-shot re-identification. Relative binary labelling performs strongest overall, especially at lower ranks. However, both binary derivations attain lower recognition rates than their counterparts for top and side camera viewpoints, supposedly less invariant to camera angle and pose.

Table 4 also includes the average rank improvement after optimisation. Interestingly, relative binary gains the least from optimisation, while still producing the best results. We

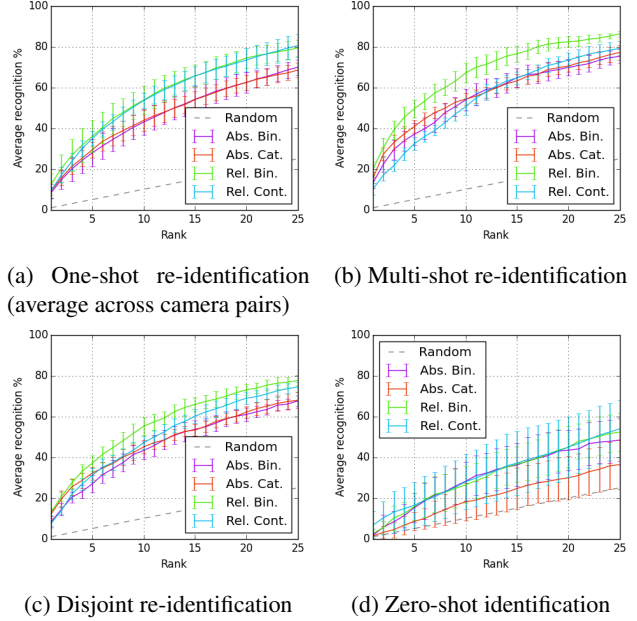


Figure 2: Recognition scenario CMCs  $\pm$ std (ranks 1-25).

Trait	Soft trait weightings $\pm$ std				
	Abs. Cat.	Abs. Bin.	Rel. Bin.	Rel. Cont.	Average
Gender	<b>2.7<math>\pm</math>0.3</b>	<b>2.6<math>\pm</math>0.6</b>	1.3 $\pm$ 0.3	<b>2.0<math>\pm</math>1.1</b>	<b>2.1<math>\pm</math>0.6</b>
Height	0.9 $\pm$ 0.4	1.0 $\pm$ 0.6	<b>2.3<math>\pm</math>1.2</b>	<b>2.2<math>\pm</math>1.7</b>	1.6 $\pm$ 1.0
Age	0.0 $\pm$ 0.6	0.4 $\pm$ 0.7	1.3 $\pm$ 1.0	0.2 $\pm$ 0.8	0.5 $\pm$ 0.8
Weight	0.5 $\pm$ 0.7	0.5 $\pm$ 0.6	0.5 $\pm$ 0.6	0.3 $\pm$ 0.2	0.5 $\pm$ 0.5
Figure	0.5 $\pm$ 0.9	0.5 $\pm$ 0.2	1.1 $\pm$ 0.9	1.5 $\pm$ 0.6	0.9 $\pm$ 0.7
Chest s.	1.1 $\pm$ 0.6	0.5 $\pm$ 0.7	1.1 $\pm$ 0.3	0.1 $\pm$ 0.5	0.7 $\pm$ 0.5
Arm t.	1.1 $\pm$ 0.5	1.5 $\pm$ 0.5	1.1 $\pm$ 0.2	0.7 $\pm$ 1.6	1.1 $\pm$ 0.7
Leg t.	-0.1 $\pm$ 0.4	0.6 $\pm$ 0.3	0.4 $\pm$ 1.0	-0.1 $\pm$ 0.8	0.2 $\pm$ 0.6
Skin c.	<b>1.8<math>\pm</math>0.3</b>	<b>1.7<math>\pm</math>0.5</b>	<b>1.5<math>\pm</math>0.4</b>	1.6 $\pm$ 0.4	<b>1.6<math>\pm</math>0.4</b>
Hair c.	<b>2.4<math>\pm</math>0.8</b>	<b>2.5<math>\pm</math>0.6</b>	<b>2.4<math>\pm</math>1.0</b>	<b>5.2<math>\pm</math>0.6</b>	<b>3.1<math>\pm</math>0.7</b>
Hair l.	<b>2.1<math>\pm</math>0.5</b>	<b>1.7<math>\pm</math>0.4</b>	<b>1.4<math>\pm</math>0.3</b>	<b>3.0<math>\pm</math>1.3</b>	<b>2.0<math>\pm</math>0.6</b>
Muscle b.	1.5 $\pm$ 0.4	1.3 $\pm$ 0.3	0.1 $\pm$ 1.9	-0.1 $\pm$ 1.0	0.7 $\pm$ 0.9

Table 5: One-shot re-identification optimised soft trait weightings (top four emboldened).

can speculate that less accurate predictions leave more room for improvement, therefore lower rank optimisation implies better overall retrieval accuracy.

We report the average soft trait weightings in Table 5, to best compare intra-label trait salience. We find *hair colour* to be most salient, likely due to accurate prediction, being highly localised in the top descriptor strip. *Gender*, *hair length* and *skin colour* follow closely, with *leg thickness*, *weight*, *age* and *muscle build* being the least discriminative traits. Intriguingly, *height* is much more prominent with relative labelling, suggesting more accurate and precise predictions as a result of better ground-truth annotations. This also reciprocates Reid et al.’s findings [24].

### 5.2. Assorted re-identification scenario results

Figure 2b-d highlights a stark performance contrast between the remaining re-identification scenarios. Relative

binary labelling again outperforms other methods in multi-shot re-identification, being able to learn a more precise model given extra data, especially at lower ranks with (Table 4b). On the other-hand, absolute categorical labelling suffers from lack of relevant training examples in the zero-shot scenario (Table 4d).

By excluding just one additional training sample per subject in disjoint re-identification, rank 10 accuracies drop on average 11.3%, showing a high reliance on similar camera viewpoints (Table 4c). The zero-shot experiment attains even poorer recognition rates, as subjects are mutually exclusive to either training or testing sets. While our retrieval methods demonstrate an ability to predict labels for unseen images of a subject, the zero-shot experiment highlights their limitations for predicting labels of entirely unseen subjects.

Zero-shot identification performance exhibits high standard deviations, causing variations in the best methods at each rank in Table 4d. However, relative continuous labelling achieves more than twice the average rank 1 accuracy in this scenario, and is the most consistent labelling mode across all scenarios.

## 6. Conclusions

We have presented a novel solution to automatically describe and identify suspects from images, given only a human description. Our approach to soft biometric retrieval and rank optimisation has been demonstrated using four semantic labelling methods, and its recognition performance has been analysed across a number of challenging re-identification experiments. We have also illustrated our publicly available soft biometric retrieval dataset and its use for simulating real-world re-identification scenarios.

Comparative ground-truth annotations have been shown to outperform categorical annotations in traditional scenarios, but are yet to fully solve the challenges presented by the zero-shot identification problem.

Our work opens many avenues for future investigations, including; joint prediction of soft traits, relative labelling techniques, low-level feature salience representations and amalgamating more advanced computer vision and machine learning algorithms.

## References

- [1] L. An, X. Chen, M. Kafai, S. Yang, and B. Bhanu. Improving person re-identification by soft biometrics based reranking. In *ICDSC*, pages 1–6. IEEE, 2013. 1, 2
- [2] D. Chen, Z. Yuan, G. Hua, N. Zheng, and J. Wang. Similarity learning on an explicit polynomial kernel feature map for person re-identification. In *CVPR*, pages 1565–1573, 2015. 1
- [3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, volume 1, pages 886–893. IEEE, 2005. 4
- [4] A. Dantcheva, P. Elia, and A. Ross. What else does your biometric data reveal? A survey on soft biometrics. *WIFS*, 2015, 04 2015. 1, 2
- [5] A. Dantcheva, C. Velardo, and A. D. et al. Bag of soft biometrics for person identification. *MTAS*, 51(2):739–777, 2011. 1, 2
- [6] I. Fogel and D. Sagi. Gabor filters as texture discriminator. *Biological cybernetics*, 61(2):103–113, 1989. 4
- [7] P. Geurts, D. Ernst, and L. Wehenkel. Extremely randomized trees. *Machine learning*, 63(1):3–42, 2006. 4
- [8] S. Gong, M. Cristani, S. Yan, and C. Loy. *Person re-identification*, volume 1. Springer, 2014. 2
- [9] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *ECCV*, pages 262–275. Springer, 2008. 2, 4
- [10] H. Han, C. Otto, X. Liu, and A. Jain. Demographic estimation from face images: Human vs. machine performance. 2014. 2
- [11] E. Jaha and M. Nixon. Soft biometrics for subject identification using clothing attributes. In *IJCB*, pages 1–6. IEEE, 2014. 2
- [12] N. Kumar, A. Berg, and P. B. et al. Attribute and simile classifiers for face verification. In *CVPR*, pages 365–372. IEEE, 2009. 2
- [13] R. Layne, T. Hospedales, and S. Gong. Attributes-based re-identification. In *Person Re-Identification*, pages 93–117. Springer, 2014. 1, 2, 4
- [14] X. Li, A. Wu, and M. C. et al. Towards more reliable matching for person re-identification. In *ISBA*, pages 1–6, March 2015. 1
- [15] C. Liu, S. Gong, C. Loy, and X. Lin. Person re-identification: What features are important? In *ECCV*, pages 391–401. Springer, 2012. 2
- [16] D. Martinho-Corbishley, M. S. Nixon, and J. N. Carter. Soft biometric recognition from comparative crowdsourced annotations. In *ICDP*, 2015. 1, 2, 3
- [17] T. Matthews, M. S. Nixon, and M. Niranjan. Enriching texture analysis with semantic data. In *CVPR*, pages 1248–1255. IEEE, 2013. 2
- [18] C. Ng, Y. Tay, and B. Goi. Recognizing human gender in computer vision: a survey. In *PRICAI*, pages 335–346. Springer, 2012. 2
- [19] M. Nixon, P. Correia, and K. N. et al. On soft biometrics. *Pattern Recognition Letters*, 2015. 1, 2
- [20] D. Parikh and K. Grauman. Relative attributes. In *ICCV*, pages 503–510. IEEE, 2011. 2, 3
- [21] H. Perlin and H. L. Extracting human attributes using a convolutional neural network approach. *Pattern Recognition Letters*, 2015. 2
- [22] S. Pizer, E. Amburn, J. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. Zimmerman, and K. Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987. 3
- [23] B. Prosser, W. Zheng, S. Gong, T. Xiang, and Q. Mary. Person re-identification by support vector ranking. In *BMVC*, volume 2, page 6, 2010. 2, 4
- [24] D. Reid and M. Nixon. Using comparative human descriptions for soft biometrics. In *TPAME*, pages 1–6. IEEE, 2011. 1, 2, 5
- [25] S. Samangooei and M. Nixon. On semantic soft-biometric labels. In *Biometric Authentication*, pages 3–15. Springer International Publishing, 2014. 1, 2, 3
- [26] 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 *BTAS*, pages 1–6. IEEE, 2008. 2, 3
- [27] R. Vezzani, D. Baltieri, and R. Cucchiara. People reidentification in surveillance and forensics: A survey. *CSUR*, 46(2):29, 2013. 2
- [28] X. Wang and R. Zhao. Person re-identification: System design and evaluation overview. In *Person Re-Identification*, pages 351–370. Springer, 2014. 2
- [29] J. Zhu, S. Liao, D. Yi, Z. Lei, and S. Li. Multi-label cnn based pedestrian attribute learning for soft biometrics. In *ICB*, pages 535–540. IEEE, 2015. 1, 2