

# From Clothing to Identity; Manual and Automatic Soft Biometrics

Emad Sami Jaha, and Mark S. Nixon

**Abstract**—Soft biometrics have increasingly attracted research interest and are often considered as major cues for identity, especially in the absence of valid traditional biometrics, as in surveillance. In everyday life, several incidents and forensic scenarios highlight the usefulness and capability of identity information that can be deduced from clothing. Semantic clothing attributes have recently been introduced as a new form of soft biometrics. Although clothing traits can be naturally described and compared by humans for operable and successful use, it is desirable to exploit computer-vision to enrich clothing descriptions with more objective and discriminative information. This allows automatic extraction and semantic description and comparison of visually detectable clothing traits in a manner similar to recognition by eyewitness statements. This study proposes a novel set of soft clothing attributes, described using small groups of high-level semantic labels, and automatically extracted using computer-vision techniques. In this way we can explore the capability of human attributes vis-a-vis those which are inferred automatically by computer-vision. Categorical and comparative soft clothing traits are derived and used for identification/re-identification either to supplement soft body traits or to be used alone. The automatically- and manually-derived soft clothing biometrics are employed in challenging invariant person retrieval. The experimental results highlight promising potential for use in various applications.

**Index Terms**—Soft biometrics, clothing attributes, computer vision, identification, relative attributes, retrieval

## I. INTRODUCTION

TRADITIONAL (hard) biometrics have proven to be a reliable and practical way to identify or authenticate a person's identity with excellent accuracy. Hence, they have received extensive use in many operational systems. However, a person's hard traits need to be correctly enrolled into a database for successful biometric use, which requires their cooperation. Furthermore, there are still challenging cases and unsolved limitations, such as low resolution and an increased distance between the camera/sensor and the captured subject, where such hard traits lose utility. Surveillance is a current example wherein the majority of hard biometrics, such as fingerprints, irises and maybe faces are impractical for identification. Soft biometrics have recently emerged as a new attribute-based form of biometrics with a high level of usability and collectability offering many advantages over hard biometrics. In contrast with hard traits, soft traits can be acquired using images/videos without a person's cooperation. Soft traits have been shown to address many problems and

overcome limitations associated with hard traits.

Several forensic scenarios and incidents highlight the usefulness and capabilities of information deduced from clothing to identify an individual. In a recent incident, a four year old girl was found dead on a Massachusetts beach in Boston in 2015, but the police have still not identified her. Therefore, the investigators tried to identify her by sharing her story via Facebook, providing a computer-generated image of her with descriptions of her face and clothing attributes, stating "A child wearing white leggings with black polka dots. Inside the bag with her remains was a zebra-print blanket".

Another recent case occurred in Nottingham in 2015, when three (masked) raiders attempted to rob a shop. As shown in Fig. 1, the police publicized surveillance footage of the incident via social media asking whether anyone could recognize these suspects by their clothing, bags, or behavior. As such real-life examples suggest that soft clothing attributes are likely to be useful, viable, and effective traits for characterizing an individual. Besides, it has been shown that soft clothing attributes are naturally correlated with each other and mutually dependent [1] or may have exclusive relations between items, which are not worn at the same time like a skirt and a dress [2]. Hence, using clothing for biometric purposes deserves more research interest than it has received.

### A. Context and Contributions

A major distinction of this work is that even though computer-vision is employed in extracting and analyzing low-level features, we further exploit these features to derive high-level soft clothing attributes. However, distinct from research on vision-based clothing attributes, we recognize people by their soft clothing attributes (as soft biometrics) rather than recognizing clothes by attributes for classification or fashion search etc. like [1-8]. Therefore, those studies without biometric interest are not related to this work. This research proposes high-level clothing descriptions that are closer to



Fig. 1. CCTV shows masked raiders (Nottingham 2015), released by police asking help to recognize them by clothing, bags, or behaviors [<https://www.youtube.com/user/nottinghampolice>].

E.S. Jaha is with the Faculty of Computing and IT, King Abdulaziz University, Jeddah, SA, (e-mail: ejaha@kau.edu.sa). M.S. Nixon is with the School of Electronics and Computer Science, the University of Southampton, Southampton, UK (e-mail: msn@ecs.soton.ac.uk).

human perception, understanding, and judgment; more generic and less detailed and also more separable, leaving smaller room for ambiguity and mismatch. The descriptions are also less sensitive to changes in illumination, viewpoint and pose; and can be used in biometrics, especially in surveillance, whilst every single trait may have its own challenges, which are related to it but not to the others [9]. For example, in describing the color attributes of a piece of clothing, some prior research used computer vision features or descriptors to categorize clothing to a limited number of basic color names (i.e. red, blue, black, etc.) serving as categorical labels [4], [10-15]. We rather use computer vision features to categorize clothing colors according to a more generic and higher level annotation describing the *Color-scheme* as ‘Warm’ like orange, and ‘Neutral’ like gray, so it is less error-prone in case of illumination variation. Another distinction of this work is that we perform automatic pairwise comparisons between subjects in a database and automatically derive comparative soft clothing traits. Comparative clothing attributes have been very rarely derived and used for biometric purposes or non-biometric like refining fashion search via comparative attributes adjusted by user feedback [16]. This research aims to be a precursor to our eventual aim, to be able to search video to find people. Therefore, it is important to show first that clothing can indeed be differentiated and used in this way.

For identification/re-identification proposes, our approach utilizes more fine-grained soft clothing attributes rather than coarse-grained attributes as in [13]. Besides, we directly use either an image or a verbal description as a query to retrieve a subject-of-interest rather than transforming semantic attributes into a searchable avatar to retrieve any matching representations as in [11]. In this research, two different groups of soft clothing biometrics are employed for the purpose of person identification and retrieval. The first group comprises manually derived traits via human-based annotations, whereas the second group comprises automatically derived traits via computer-vision techniques. We focus more on the new automatic vision-based soft clothing biometrics in terms of explanation, analysis, and discussion, since we have already paid attention to the human-based soft clothing biometrics that are explained in detail [17], extensively analyzed [18], and discussed for use in challenging scenarios [19].

The novelty of this paper is how to automatically extract and annotate clothing for producing a biometric signature of soft clothing traits for an individual. We show how individuals can be identified and retrieved from the database by using either query images or verbal descriptions. The main contributions of this research are:

- introducing a set of automatic soft clothing attributes to bridge the semantic gap between vision-based low-level features and high-level clothing annotations for person identification and retrieval;
- a new approach for automatic soft clothing attribute annotation via a combination of established vision-based techniques, exploiting the precision of computer vision and emulating the tolerance of human vision;

- automatically deriving categorical and comparative forms of soft clothing traits from automatic annotation and comparison data; and
- viewpoint invariant subject retrieval using automatic and manual soft clothing biometrics in isolation and in fusion with other body-based soft biometrics.

### B. *Soft Biometrics and Identity*

Recently, several soft biometric traits have been introduced to offer a new solution to some of the problems of biometric systems. Soft body and face biometrics have been attracting increasing research interest and are often considered as major cues for identifying individuals [20-25], especially in the absence of valid traditional hard biometrics. The basic approach uses human-vision wherein labelers describe human body features using understandable labels and measurements, which in turn allow for recognition and retrieval using only verbal descriptions as the sole query [23], [26]. The features also allow prediction of other measurements as they have been observed to be correlated [27]. Indeed, soft traits are not unique to an individual but a discriminative biometric signature can be designed from their aggregation. Since verbal identification can be used to retrieve subjects already enrolled in a database [24], it could be also extended, in a more challenging application, for retrieval from video footage [23]. The capability of verbal retrieval from images and videos can pave the way for applications that can search surveillance data of a crime scene to match people to potential suspects described verbally by eyewitnesses.

Soft biometric databases based on categorical labels can be incorporated with other biometrics to enhance recognition, such as integrating soft body traits with a gait signature [24], and using soft facial traits along with other (hard) facial traits [28]. Nevertheless, soft comparative labels have been demonstrated to be more successful in representing the slight differences between people in bodily descriptions [23]. Facial marks, for instance, can be automatically detected and described to be used as micro-soft traits to supplement primary facial features for improved face recognition and fast retrieval; besides, they may enable matching with low resolution or partial images [29]. Measured facial information might be useful for gender prediction [30] and many system issues and challenges could arise when soft facial traits are used at a distance [20]. For surveillance purposes, different forms of soft biometrics take place in a variety of applications and scenarios [11], [12], [15], [23], [25], where many of those soft traits could be distinguished even at a distance and then be fused with biometric recognition systems for the sake of improving the overall recognition, especially with poor quality surveillance videos [25].

Clothes are a predominant visible characteristic of a person’s appearance [18]. However, clothing has rarely been adopted to represent soft biometric traits for an individual and has been considered unlikely to be a clue to identity [18]. Clothing can reflect some cues regarding social status, lifestyle and cultural affiliation. In addition, clothing encodes more information about an individual, beyond just their visual

appearance [3]. People choose and buy clothing via semantic attributes. Defining and utilizing a list of clothing attributes for various objectives has been the concern of several non-biometric [1-8] and few biometric [18], [31], [32] studies.

There are few other research studies associated with using clothing for biometric purposes, while a number of most recent approaches were reviewed in [33]. The majority of existing research employs computer-vision algorithms and machine learning techniques to extract and use clothing descriptions in applications including: online person recognition [28], [34]; soft attributes for re-identification [13], [31], [35] along with person detection [11], [12] and tracking [36], [37] or attribute-based people search [11], [13]; detecting and analyzing semantic descriptions (labels) of clothing colors and types to supplement other bodily and soft facial attributes in automatic search and retrieval [15]; and utilizing some clothing attributes like color [38] [39] and style to improve observation and retrieval at a distance in surveillance environments [26]. It will be difficult to analyze clothing in some surveillance images, given poor quality and low resolution, while human-vision analysis offers supportive or alternative solutions [17]. Clothing is innately more efficient in short term id/re-id as people might change their clothes [19]. Even with images captured on different days, there remains sufficient information to compare and establish identity, since clothes are often re-worn or a particular individual may prefer a specific clothing style or color [40]. Clothing descriptions like indicative colors and decorations could be utilized to supplement other behavioral biometrics, like human motion patterns, hence they can form a biometric fingerprint that serves as a person's identifier [34] or other than identification, to recognize gender via region-based clothing info in the case of insufficient face specification [41].

People can be unique regarding their clothing and many people often wear similar clothing from day to day, or a certain clothing style [17]. Soft clothing traits are a new form of soft biometrics that can be associated with biometric signatures to achieve successful subject retrieval [18]. This motivates more interest in the latent ability of clothing information in subject retrieval and identification. Learning reliable biometric traits is required for realistic scenarios, such as after a change of viewpoint and partial occlusion [42]. In such scenarios, even some soft biometrics may likely be more vulnerable, especially to annotation subjectivity and missing information mostly caused by occlusion [18]. Viewpoint invariance is a challenging problem that has been considered in most biometric modalities. Subject retrieval is deemed as viewpoint invariant, if it remains invariant to any angle from which a subject is likely to be seen [43], such as front and side views. Soft clothing attributes can be naturally described by humans for operable and successful use in identification and re-identification. It is still desirable to enrich clothing descriptions by exploiting computer-vision to extract and semantically describe some visually detectable soft attributes.

The rest of this paper provides in Section II a brief description of the manual soft clothing attributes and human-based annotations. Section III explores the automatic soft

clothing attributes and the methodology used for vision-based automatic extraction and annotation. Section IV presents the consequent soft clothing biometrics and their galleries of clothing descriptions. Detailed analysis of soft clothing data is given in Section V. Experiments, results, and conclusions are provided in Section VI and VII.

## II. HUMAN-BASED SOFT CLOTHING ATTRIBUTE ANNOTATION

For further analysis and performance comparison we use our previously proposed manual soft clothing biometrics for human identification [17] and subject retrieval [18], [19] to be evaluated and integrated with our new automatic soft clothing biometrics in Section III. These soft biometrics were derived from human-based manual labeling for a proposed set of 21 soft clothing attributes, including: **1. Head clothing category**, **2. Head coverage**, **3. Face covered**, **4. Hat**, **5. Upper body clothing category**, **6. Neckline shape**, **7. Neckline size**, **8. Sleeve length**, **9. Lower body clothing category**, **10. Shape of lower clothing**, **11. Leg length of lower clothing**, **12. Belt presence**, **13. Shoes category**, **14. Heel level**, **15. Attached object category**, **16. Bag (size)**, **17. Gun**, **18. Object in hand**, **19. Gloves**, **20. Style category**, **21. Tattoos**. These are presented and discussed in detail with their descriptive labels in [17]. Each of these soft attributes is described using a set of appropriate categorical labels; for example, (*Sleeve length*: 'Very short', 'Short', 'Medium', 'Longer', 'Much Longer').

Seven of the 21 soft attributes (shown in bold) are derived using comparison such as (*Neckline size*: 'Much smaller', 'Smaller', 'Same', 'Larger', 'Much larger'). For non-relative soft attributes, each categorical label is assigned an integer value to represent the textual expression of the label. For better representation of relative soft attributes via categorical and comparative labels, we define bipolar scales in a way inspired from an early analysis that characterized human traits for whole-body descriptions [44]. Each of the seven relative soft attributes is formed as a categorical/comparative bipolar of a five-point scale. To obtain clothing annotations, we designed a website for online data acquisition. Thus, all front- and side-view samples of the 128 subjects were labeled by a group of 27 annotators and each subject was annotated and compared by multiple annotators, providing 21 categorical labels describing a single subject independently and seven comparative labels describing them with respect to another subject. From the available collected comparisons, we enrich comparative label data by inferring all possible comparisons, such that if subject  $i$  was compared with subjects  $j$  and  $k$ , we can infer the unavailable comparison between  $j$  and  $k$ .

## III. VISION-BASED SOFT CLOTHING ATTRIBUTE EXTRACTION

### A. Automatic Soft Clothing Attributes

In the majority of existing vision-based research, soft clothing attributes are suggested by either mining web fashion data [14] or manual designs allowing to suggest more visually separable soft attributes to be more suited for biometric use [35]. Unlike most existing vision-based approaches concerning clothing attributes for fashion search/retrieval or

TABLE I  
 AUTOMATIC SOFT CLOTHING ATTRIBUTES AND CORRESPONDING LABELS

Body Part	Soft Clothing Attribute	Categorical Labels	Comparative Labels
<b>Overall body</b>	<b>A1. Overall Skin exposure</b>	[Very low, Low, Medium, High, Very high]	[Much lower, Lower, Same, Higher, Much higher]
	A2. Overall Clothing season	[Cold, Neutral, Hot]	
	<b>A3. Upper vs Lower Contrast</b>	[Very low, Low, Medium, High, Very high]	[Much lower, Lower, Same, Higher, Much higher]
	A4. Overall Color-scheme	[Cool, Warm, Neutral, Mixed]	
<b>Upper body</b>	<b>A5. Upper Skin exposure</b>	[Very low, Low, Medium, High, Very high]	[Much lower, Lower, Same, Higher, Much higher]
	A6. Upper Clothing season	[Cold, Neutral, Hot]	
	<b>A7. Upper Brightness</b>	[Very dark, Dark, Average, Light, Very light]	[ Much darker, Darker, Same, Lighter, Much lighter]
	A8. Upper Color-scheme	[Cool, Warm, Neutral]	
	<b>A9. Upper Dominant color</b>	[Single, Dual, Multiple]	[Less, Same, More]
	<b>A10. Upper Pattern</b>	[None, Simple, complex]	[Much simpler, Simpler, Same, More complex, Much more complex]
<b>Lower body</b>	<b>A11. Lower Skin exposure</b>	[Very low, Low, Medium, High, Very high]	[Much lower, Lower, Same, Higher, Much higher]
	A12. Lower Clothing season	[Cold, Neutral, Hot]	
	<b>A13. Lower Brightness</b>	[Very dark, Dark, Average, Light, Very light]	[ Much darker, Darker, Same, Lighter, Much lighter]
	A14. Lower Color-scheme	[Cool, Warm, Neutral]	
	<b>A15. Lower Dominant color</b>	[Single, Dual, Multiple]	[Less, Same, More]
	<b>A16. Lower Pattern</b>	[None, Simple, complex]	[Much simpler, Simpler, Same, More complex, Much more complex]
	A17. Footwear category	[Closed toed, Open toed]	

clothes recognition/classification, this research proposes a novel set of soft clothing attributes for human identification and retrieval, each soft attribute is full-automatically derived and described at a high-level using a small group of semantic *categorical/comparative* labels for biometric purposes.

It is noteworthy that a categorical label describes a relative or non-relative soft attribute with respect to a single subject independently, whereas a comparative label describes only the relative soft attribute of a single subject in comparison with another subject. Hence, non-relative soft attributes, which may be binary or multi-class soft attributes, can be annotated using only categorical (absolute) labels such as (*Footwear category*: ‘Closed toed’, ‘Open toed’). On the other hand, relative soft attributes, which are also comparable, can be either annotated using categorical labels representing the degree-of-strength of the relative soft attribute, or can be annotated using comparative label reflecting the degree-of-comparison of the relative soft attribute. Out of the 17 soft clothing attributes listed in Table 1 only ten soft attributes, shown in bold, are relative and can be described by a categorical form of labeling (such as *Upper Skin exposure*: ‘Very low’, ‘Low’, ‘Medium’, ‘High’, ‘Much higher’), and also suited for comparison to be described using a comparative form of labeling (such that *Lower Brightness*: ‘Much darker’, ‘Darker’, ‘Same’, ‘Bright’, ‘Much brighter’).

Likewise, as described in Section II, for non-relative soft attributes, each categorical label is coded by an integer value representing the label. For relative soft attributes, each of the ten relative soft attributes is formed as a categorical or comparative bipolar of a three- to five-point scale; where each point is assigned a suitable value from a set of ordered label codes ranging from 1 to 3 or 1 to 5. Note that, for the relative soft attributes of dominant colors and clothing patterns we exceptionally use a bipolar three-point scale for categorical and comparative forms of (A9 and A15) and for categorical form of (A10 and A16) to avoid ambiguity and overestimation of such soft attributes. Table I shows the proposed set of soft attributes describing clothing with regards to three different aspects: *Overall body*, *Upper body*, and *Lower body*. From computer-vision perspective, these soft attributes can be categorized into three groups: The first group comprising (A1,

A2, A5, A6, A11, A12, and A17) associated with clothing appearance, shape, and style; the second group including (A3, A4, A7, A8, A9, A13, A14, and A15) semantically describing clothing colors, brightness, and contrast; and the third group of (A10 and A16), characterizing clothing patterns.

### B. Automatic Soft Clothing Attribute Annotation

Based on computer-vision, we propose a novel approach for automatic soft clothing attribute extraction and annotation. For better description of a soft attribute, this approach is designed in such a way as to integrate both the judgment strategies of computer-vision and human-vision. In other words, a way exploiting the precision and consistency of the machine, as well as emulating the (useful) tolerance and approximation of the human, leading to more harmonious judgments suited to the nature of soft biometrics. This approach starts with the input of an image of a subject and ends with a list of 17 categorical soft labels describing the subject’s clothing to be fully or partially used as soft clothing traits for biometric retrieval, as shown in Fig. 3. The process of this approach comprises four main phases: preprocessing, analyzing appearance, analyzing colors, and analyzing patterns. Fig. 2 illustrates an overview of our experimental framework for automatic soft clothing attributes extraction and labelling.

#### 1) Preprocessing

The proposed approach is designed to enforce soft clothing attribute extraction on subject images with different viewpoints (i.e. front and side). Therefore, a combination of body and face detection is considered for better human

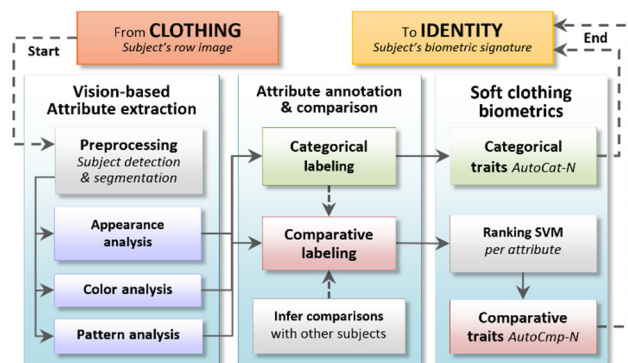


Fig. 2. Overview of automatic soft clothing attribute extraction &amp; labeling.

detection results. This enables a more accurate and reliable detection of the subject in an image.

Holistic human body detection is achieved using a deformable model-based detector [45], and a cascade classifier based on Haar-like features is used for face detection [46]. In this phase we aim to detect both body and face when possible, or to detect either of them and estimate the other if undetectable. In both cases, reassessment and realignment are applied exploiting the intersection and relative mutual positioning between the two estimated bounding-boxes of body and face, with a higher priority placed on the detected face, resulting in enhanced detection (see Fig. 3). Hence, the body box may be realigned to the face box and vice versa.

Afterwards, the final estimated region containing the face/head (if no detectable face) is used to initialize a starting region of a foreground highlighting method [47], which employs the Grabcut technique for foreground object extraction [48] by learning foreground and background color models that initially suggest sub-regions where the person is likely to be present or absent. This method is used here to segment the subject's body as foreground pixels and to discard most of the background pixels. Since the resulting segmentation output is likely to contain some background clutter, we design and use new color models, shown in Fig. 4, to attain accurate full body segmentation. These color models are generally designed with similar concepts as in [47] but with three main characteristics.

First, the color models are more conservative in extracting shape and details of the full appearance of the person, including the lower body and feet regions, which are also to be considered in the foreground model as likely candidate

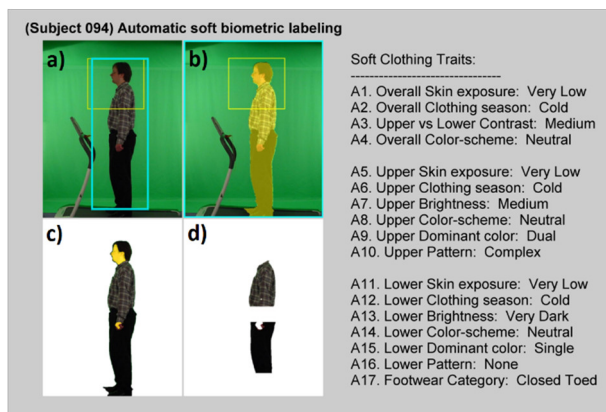


Fig. 3. A summary screen of a set of all 17 categorical soft clothing attributes automatically extracted and labeled to describe a side viewpoint subject image. Sub-images a, b, c and d show some applied steps of automatic vision-based image processing, including detection, analysis, segmentation, etc.

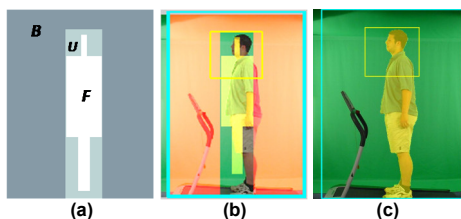


Fig. 4. (a) Color models used to initialize the Grabcut person extractor, (b) Resized and fitted color models initializing a foreground highlighting method in a tested subject image. (c) Resulting highlighting for subject segmentation.

regions. Second, the color models embody three sub-regions and none of these sub-regions are clamped-as-foreground known as ( $F_c$ ), so Grabcut is able to reset any sub-region. Such that, if  $\mathcal{R}$  is the enlarged region of the body detection window centered by the face box.  $\mathcal{R}$  is divided into three sub-regions:  $F$  mostly representing foreground but maybe containing some background (no  $F_c$  sub-region to be enforced as definite part of the body),  $B$  mostly representing background, and  $U$  being a neutral region equally free to form either foreground or background. Third, the initial foreground color model  $F$  is rescaled to mostly fit within the person's body captured from the side viewpoint as well as the front.

Note that gamma correction is used to enhance the input images before segmentation. In this way, we obtain more effective body segmentation suited to the context and objectives of this research, with much less background clutter. Further processing is applied to refine and improve the segmentation output. Particularly the noise pixels removal, which may appear around the feet region due to the shadow effect, or may exist as undesirable fragments of background or other objects. Subsequently, to allow the three-part-based clothing analysis shown in Table I, the full body segment is processed as a whole to derive overall soft clothing attributes, and it is further segmented into two parts to derive corresponding upper and lower soft clothing attributes. The upper body sub-region is defined relative to the face box to be around 42% of the full body including the face box, while the lower body sub-region is expected to embrace the rest of the full body region. Note that these sub-regions are in accord with the human body segment properties proposed in [49].

## 2) Analyzing Appearance

Skin exposure soft attributes (A1, A5, and A11) and clothing season soft attributes (A2, A6, and A12) are deduced for the three body parts overall, upper, and lower, whereas 'Footwear category' (A17) is deduced as a separate soft attribute attached to the lower body part.

Skin is detected by exploiting four methods. A pre-trained model-based skin detector [50] is used to compute a pixel-level skin likelihood probability. Then we apply three skin tone detectors, in three color spaces, HSV, YCbCr, and RGB, by defining the ranges of color values for each channel, which are likely to represent the variety of skin colors. The model-based skin detector is sensitive to shaded regions and small details in body-level detection (such as eyes, eyebrows, and mouth) producing noisy and distorted skin segmentation. Thus, a fusion approach is adopted to achieve better skin segmentation suited to the body-level.

The model-based detector is combined with each of the three skin-color-based detectors. The YCbCr detector was found to be the most effective and coherent in detecting most skin color tones and it might neglect some small details and edges, which are likely to be detected by the model-based detector. Therefore, the combination of the model-based and the YCbCr detectors is used as a default candidate skin mask. This mask is then validated in terms of noise and distortion, and compared with another two masks, respectively derived by the other two combinations of the model-based detector

(with the HSV detector) and (with the RGB detector). As such, a skin mask with the largest connected components and least noise fragments is nominated as the final inferred skin segmentation. Note that we often obtain an accurate skin detection, as the detection is applied to the segmented subject in image with no confusing background.

Skin exposure description, as a relative soft clothing attribute, can implicitly reflect further information about the type/style of worn clothing. In other words, the type/style of clothing intuitively causes an expected amount of human-skin to be revealed. Consequently, this observation can be exploited to further describe a conforming seasonal category (e.g. hot, cold, and neutral), which such clothing are likely to belong to. We utilize this knowledge to implement a systematic auto-labeling, which is also adaptive to the change in viewpoint from front to side.

In order to annotate ‘Skin exposure’ and ‘Clothing season’ for overall, upper, and lower body clothing, they are mainly inferred from derived vision-based relative measurements for each of the three body parts. These relative measurements are computed with respect to the proportion of exposed skin, and then compared with a pre-defined scale with five/three ranges of the five/three categorical labels, leading to the assignment of the best representative label for each soft attribute.

Subsequently, a sliding window is used to detect and examine the region containing the feet to annotate the ‘Footwear category’. This is performed by utilizing the skin detection in the feet zone to investigate whether the footwear is closed toed, covering the skin of the entire feet, or open toed, exposing some of skin. Eventually the detected skin is omitted from all body segments and these segments are refined to prepare a pure clothing representation to be passed and used in the next clothing colors and patterns analysis.

### 3) Analyzing Colors

Clothing colors are analyzed from four different perspectives to extract a variety of soft attributes describing: brightness of upper and lower garment pieces (A7, and A13); overall contrast between upper and lower clothing (A3); quantitative description of dominant colors of upper and lower parts (A9, and A15); and color-scheme category of overall, upper, and lower clothing (A4, A8, and A14). The RGB upper/lower clothing segment is converted to 256 indexed color space and then approximated to 64 indexed colors using a minimum variance quantization, to be converted further to grayscale color space for brightness analysis. This process is useful for minimizing clothing wrinkles and shadow effects. We design and use a five-scale color-map model classifying all possible grayscales (from 0 to 255) into five groups of  $K=5$  brightness levels. Each level is interpreted to a suitable categorical label. As such, for each tested segment a maximum of 32 most frequent pixel colors are detected and classified, each into one of the five brightness groups (or levels) as applicable. Then these  $K$  color groups are used along with a voting formulation  $V$  defined in (1), to nominate the most appropriate label  $L_k$  that earns the highest vote.

$$V(L_k) = \frac{1}{M} \sum_i^{N_k} \sum_j^M v(p_j, c_i) \quad (1)$$

where  $N_k$  is the number of colors in  $k^{\text{th}}$  group represented by the label  $L_k$ .  $M$  is the total number of pixels in the clothing segment, and  $1/M$  is used to deduce a normalized relative measurement.  $v$  is the voting value assigned for  $j^{\text{th}}$  pixel  $p_j$  with respect to color  $c_i$ , such that it equals to 1 if  $c_i$  is the current color of  $p_j$  and otherwise it is set to 0, as in (2).

$$v(p_j, c_i) = \begin{cases} 1 & \text{if } p_j \text{ is colored with } c_i \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

However, in some exceptional cases, if there are two or more labels competing with (almost) similar vote totals, a middle label of the average color is nominated instead. For example, a black and white striped T-shirt is annotated as ‘Average’ in brightness. The extracted upper and lower brightness soft attributes are utilized further to describe the contrast of upper versus the lower clothing brightness.

For color-scheme detection and annotation, each RGB clothing segment is converted to HSV, then reduced to a maximum of 64 indexed colors. A color-scheme model is designed and used with the Euclidean distance matching method to categorize each color into one of three color-scheme categories ‘Warm’, ‘Cool’, and ‘Neutral’ as shown in Fig. 5. Since the three-valued color form (such as RGB or HSV) is unsuited to use as a conventional vector to compute the Euclidean distance, and is unreliable for discriminative kNN classification due to the confusion between different colors resulting in the same (squared average-sum) distance. Thus, all matched colors in the model and in the clothing segment are properly mapped to unique color codes to avoid confusion and mismatch.

Hence, a set of  $K$  (one to three) labeled color groups is used with the same multi-case voting formulation described in (1) and (2) to nominate the winning label  $L_k$ . Our color-scheme model classifies all grayscale colors as ‘Neutral’. Moreover, if two different color-schemes (almost) equally exist in the upper or lower segment we also assign the label ‘Neutral’. For the ‘Overall color-scheme’ annotation we add the label ‘Mixed’, reflecting that upper and lower clothing have two different color-schemes, as they are two separate (i.e. upper and lower) clothing pieces, obviously representing a new notable clothing style. Eventually the dominant colors for each of the upper and lower clothing segments are detected and semantically annotated. This is performed via a method for color segmentation using a k-means clustering algorithm [51], which utilizes the information of color histograms to select the dominant colors from the input image to be initialized as the k-means for the clustering process.

Hence, each RGB clothing segment is converted to (four to nine) indexed colors then converted to HSV form. We initially specify  $K=4$  to be the maximum number of means of clusters in a clothing image. The formulation in (3), with two cases defined in (4), is used to deduce the final integer value  $K'$  where  $1 \leq K' \leq 3$  representing the number of estimated



Fig. 5. RGB color representation for the color-scheme model.

clusters (colors) in the image to be either one, two, or three, which implicitly describes the number of dominant colors in clothing, in a high-level annotation to be either ‘Single’, ‘Dual’, or ‘Multiple’. Note that  $K'=3$  with the corresponding label ‘Multiple’ reflects that a piece of clothing has three-plus dominant colors. Since the background, in a clothing segment, has the same color and a large number of pixels, it is always accounted for as one dominant cluster. Hence the total of  $K'$  is subtracted by one to exclude the background cluster/color such that:

$$K' = \left( \sum_i^K \omega(u_i) \right) - 1 \quad (3)$$

where

$$\omega(u_i) = \begin{cases} 1 & \text{if } \frac{|u_i|}{M} \geq \epsilon \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where  $\omega$  is the weighting function of  $i^{\text{th}}$  cluster  $u_i$  assigning the value 1 and considering  $u_i$  as a dominant cluster if the proportion of pixels belonging to this cluster is greater than or equal to the defined threshold  $\epsilon$  of the minimum proportion of a dominant cluster.  $|u_i|$  is the size of the cluster divided by  $M$ , the total number of pixels in the clothing segment.

#### 4) Analyzing Patterns

A clothing segment is processed to extract and analyze visual clothing patterns in terms of simplicity and complexity, leading to a high-level description of the soft attributes (A10 and A16) of upper and lower clothing patterns. As such, each clothing segment image is cropped with a minimum rectangular border surrounding the clothing region edges. Gamma correction is applied to the color image before it is converted to grayscale. We use uniform Local Binary Patterns (ULBP) [52] as in (5) and (6). Here the uniform pattern detection accompanies a basic LBP operator defined with a circular symmetric neighbor set of  $P=8$  members on a circle of radius  $R=1$ , where  $riu2$  denotes the use of rotation invariant uniform patterns that have a value of at most 2. The uniform method is adopted to be  $U \leq 2$  transitions, namely adopting 58 values out of all possible integer grayscale values, ranging from 0 to 255, where all the non-uniform patterns  $U > 2$  are grouped as a miscellaneous pattern and assigned the value 0.

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{if } U(LBP_{P,R}) \leq 2 \\ P + 1 & \text{otherwise} \end{cases} \quad (5)$$

where

$$U(LBP_{P,R}) = |s(g_{p-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \quad (6)$$

Hence for each neighboring gray value ( $g_p$ ), a local neighborhood is thresholded at the gray value of the center pixel ( $g_c$ ) into a binary pattern by computing the gray value difference  $x = (g_p - g_c)$ , to then set the 0/1 sign of difference  $s(x)$ , where  $x$  is set to 1 only if greater than a pre-defined threshold  $\epsilon$  (rather than 0 difference as in [52]). As such, we reduce the sensitivity of the LBP operator by considering only the signs of gray value differences that are larger than  $\epsilon$  to be the minimum detectable gray-difference between two compared pixels  $g_p$  and  $g_c$  as stated in (7).

$$s(g_p - g_c) = s(x) = \begin{cases} 1 & \text{if } x > \epsilon \\ 0 & \text{if } x \leq \epsilon \end{cases} \quad (7)$$

Here the annotation procedure of clothing patterns includes the notion that a clothing pattern is more complex if it has more details and more colors with a high variance, resulting in more ULBP density. Clothing edges are very likely to be detected as distinct patterns and major parts to represent the ULBP, which is undesirable here as we are concerned with embedded patterns of pieces of clothing, not the color transitions between clothing and background so edge points detected by Prewitt are removed. Thus, we compute the pixel-sum by counting all non-zero pixels of the produced ULBP image. A proper label is assigned based on a pre-trained scale model defining the numeral range of the three labels (‘None’, ‘Simple’, and ‘Complex’).

#### C. Automatic Soft Clothing Attribute Comparison

Beyond the automatic soft attribute annotation, our approach aims to provide an automatic soft attribute comparison resulting in further comparative soft biometric information. For each of the ten comparable relative soft attributes, we first generate all  $\binom{128}{2}$  possible combinations of pairwise subject comparisons, with respect to a single soft attribute. In each comparison, a suitable comparative label is inferred to best describe the difference in the *strength* of that soft attribute between two compared subjects. As described in Section III-A, each relative soft attribute is formulated to represent the degree-of-comparison of a soft attribute with a bipolar five-point scale ranging from 1 to 5. As such, the lowest label (or degree) such as ‘Much Simpler’ is assigned the code 1 whereas the highest label such as ‘Much Complex’ is assigned the code 5. While the similarity label ‘Same’ is always coded as 3 to reflect the middle, neutral, point.

Automatic comparative annotation is achieved via a major- and a sub-process as in (8) and (9). With respect to a single compared soft attribute  $a$ , for an ordered paired comparison between two subjects  $i$  and  $j$ , we *compare* their corresponding categorical label codes, denoted  $L_i$  and  $L_j$ , which were earlier extracted and categorically annotated, as in Section III-B.

$$compare_a(L_i, L_j) = \begin{cases} 1 & \text{if } L_i - L_j \leq -2 \\ 2 & \text{if } L_i - L_j = -1 \\ sim_a(X_i, X_j) & \text{if } L_i - L_j = 0 \\ 4 & \text{if } L_i - L_j = 1 \\ 5 & \text{if } L_i - L_j \geq 2 \end{cases} \quad (8)$$

where

$$sim_a(X_i, X_j) = \begin{cases} 2 & \text{if } X_i < X_j \\ 3 & \text{if } X_i = X_j \\ 4 & \text{if } X_i > X_j \end{cases} \quad (9)$$

By using (8), a suitable label is assigned in all cases but one if  $L_i = L_j$ . So in this case, when  $L_i - L_j = 0$ , a further check is needed to ensure that the similarity  $sim_a$  is strong enough and to test if there is a slight difference that is undetectable by (8). The similarity is confirmed at a low level by comparing the relative feature values  $X_i$  and  $X_j$ , which represent the strength of presence of the soft attribute  $a$  for subjects  $i$  and  $j$  respectively. Thereby, the automatic comparison task allows to emulate the human perception ability to detect the slight visible difference between two subjects with respect to a particular soft attribute.

### 1) Ranking Subjects by Relative Soft Attribute

In order to deduce a set of comparative clothing traits for each subject in database, a *relative measurement* should be derived for each relative soft attribute. The relative measurement represents a comparative soft trait, reflecting the degree of presence of a single comparable soft attribute. Since each relative measurement is inspired by relevant comparisons between subjects, it is supposed to provide a measure of a particular soft attribute in relation to the rest of the subjects in the database. These relative measurements can be derived by using ranking methods to arrange a list of ordered subjects with respect to a single soft attribute. While all comparisons between subjects according to that soft attribute are used as rules to enforce ordering, and to adjust a relative measurement for each subject. Here, to achieve order and to derive the desired relative measurements, a soft-margin Ranking SVM method [53] is used, along with a supporting formulation of similarity constraints [54], as shown in (11). This is done to apply a pairwise technique based on learning a ranking function per soft attribute. Such learned ranking functions can be used not only to perceive the relative strength of soft attributes in a training sample, but also to predict the relative strength of a new test sample. As shown in (10), for a set of soft attributes  $A$ , a ranking linear function  $r_a$  is learned for each soft attribute  $a$  such that:

$$r_a(x_i) = w_a^T x_i \quad (10)$$

where  $w_a$  is the coefficient of the ranking function  $r_a$  and  $x_i$  is a feature vector of soft attributes of the  $i^{\text{th}}$  subject being ranked. A set of comparisons is rearranged into two groups to represent the pairwise relative constraints required to learn a ranking function. The first group consists of a set of dissimilarity comparisons  $D_a$  of ordered pairs so that  $(i, j) \in D_a \Rightarrow i > j$ , namely subject  $i$ , has a stronger presence of soft attribute  $a$  than  $j$ , whereas the second group comprises a set of similarity comparisons  $S_a$  of non-ordered pairs so that  $(i, j) \in S_a \Rightarrow i \sim j$ , namely  $i$  and  $j$ , have a similar presence of soft attribute  $a$ .  $D_a$  and  $S_a$  sets are then utilized to derive the  $w_a$  coefficients of  $r_a$  according to the following formulation:

$$\begin{aligned} & \text{minimize} \quad \left( \frac{1}{2} \|w_a^T\|^2 + C \sum \xi_{ij}^2 \right) \\ & \text{subject to} \quad w_a^T(x_i - x_j) \geq 1 - \xi_{ij}; \quad \forall (i, j) \in D_a \\ & \quad \quad \quad |w_a^T(x_i - x_j)| \leq \xi_{ij}; \quad \forall (i, j) \in S_a \\ & \quad \quad \quad \xi_{ij} \geq 0 \end{aligned} \quad (11)$$

The degree of misclassification is measured by  $\xi_{ij}$  and the trade-off between maximizing the margin and minimizing the error (i.e. satisfying constraints) is denoted as  $C$ . The resulting optimal  $w_a$  function is able to enforce (explicitly) a desirable ordering for all training samples, with respect to  $a$ . A feature vector  $x_i$  is mapped by (10) to a corresponding feature vector comprising a number of real-value relative measurements.

## IV. SOFT CLOTHING BIOMETRICS

### A. Categorical Soft Clothing Traits

The vision-based clothing annotations are used to compose for each subject two categorical feature vectors of automatic clothing descriptions. The first feature vector is composed

using the full set of 17 automatic categorical clothing traits describing a single subject. Then we apply feature subset selection leading us to nominate 14 automatic categorical clothing traits. These 14 traits are the most effective and discriminative traits (the top 14 listed in order in Fig. 7-a and Table III-a), which are selected based on ANOVA (described in Section V-C), and a number of similarly structured feature vectors for all subjects are gathered to construct the first categorical gallery of 14 automatic clothing traits (referred to as *AutoCat-14*).

Likewise, the human-based clothing annotations are used to form the second feature vector of manual clothing descriptions. This feature vector is composed per subject using the full set of 21 manual categorical clothing traits; each trait is represented by a normalized average-label deduced from a set of labels provided by multiple annotators describing the same trait of the same subject. Then we select a feature subset of the top six correlated manual traits (2, 4, 1, 9, 8, and 11, as in Section II), which are also examined via ANOVA and found to be a subset comprising six of the most effective and successful traits in subject discrimination. By grouping all these feature vectors, we build the second categorical gallery of six manual clothing traits (referred to as *Cat-6*).

### B. Comparative Soft Clothing Traits

We use the ranking SVM method, explained in Section III-C-1, to derive a comparative form of the soft clothing biometrics. Hence, in both automatic and manual versions of the clothing description data, the available collected and inferred comparative labels are used to derive the desired relative measurements forming the comparative clothing traits.

The automatic categorical feature vectors of all subjects, where each comprises only ten values describing the ten automatic relative traits, are used as a training dataset to learn ten optimal ranking functions for the ten automatic relative soft attributes (A1, A3, A5, A7, A9, A10, A11, A13, A15, A16), shown in bold in Table I. The coefficient  $w$  is derived by using the formulation in (11). The desirable per soft attribute ordering of all subjects is derived by using  $w$ . Then by (10), each  $w$  is used to map each categorical (relative) 10-value feature vector to a corresponding comparative feature vector of ten relative measurements (i.e. comparative traits) describing a single subject. Then a feature subset selection via ANOVA is also applied, here resulting in the nomination of the five most effective and discriminative comparative traits (the top five are listed in order in Fig. 7-b and Table III-b). Consequently, all obtained similarly structured comparative feature vectors are gathered to build the first comparative gallery of five automatic clothing traits (referred to as *AutoCmp-5*). Similarly, all manual categorical feature vectors of all subjects are used as a training dataset to learn seven optimal ranking functions for the seven relative soft attributes (2, 3, 7, 8, 11, 14, and 16), shown in bold in Section II. Using (11), we derive the coefficient  $w$  toward enforcing ordering per soft attribute for all subjects, and with (10) we map each 7-value categorical feature vector to a corresponding comparative feature vector of seven manual comparative traits



TABLE II  
THE OBTAINED AUTOMATIC/MANUAL SOFT CLOTHING DATA

Manual data summary	Collected	Inferred	Total
Total user annotations	444	N/A	444
Total user comparisons	317	556	873
Total attribute annotations	9324	N/A	9324
Total attribute comparisons	2219	3892	6111
Automatic data summary	Collected	Inferred	Total
Total subject auto-annotations	N/A	256	256
Total subject auto-comparisons	N/A	8385	8385
Total attribute auto-annotations	N/A	4352	4352
Total attribute auto-comparisons	N/A	83850	83850

describing a single subject. Thereby, all such mapped feature vectors are gathered to compose the second comparative gallery of seven manual clothing traits (referred to as *Cmp*).

### V. SOFT CLOTHING DATA ANALYSIS

#### A. Soft Clothing Description Data

In this research we use a standard data subset of human full body images comprising the front- and side-view samples of 128 subjects from the Soton Gait database [55]. Note that it appears sufficiently challenging for the current initial study that about 90% of the subjects in the database wear similar clothing (jeans, T-shirt, etc.). We use this image subset for automatic and manual soft clothing attribute annotations, leading to a new database of soft clothing descriptions for biometric identification and retrieval. All images and labels are available at [http://users.ecs.soton.ac.uk/msn/soft\\_labels/](http://users.ecs.soton.ac.uk/msn/soft_labels/). Table II shows a summary of all collected or inferred automatic and manual soft clothing annotations or comparisons. Soft body biometric data is already available for the same subjects in the database, which was previously collected and used [23], [24]. These soft body biometrics are used here as a baseline for performance comparison with our soft clothing biometrics and are also used as complementary traits supplemented by soft clothing traits.

#### B. Correlations between Automatic Soft Attributes

For better understanding of the usability and operability of the proposed soft clothing traits, the correlations between soft clothing attributes are investigated. Fig. 6 shows the correlation matrix of all automatic soft clothing attributes (see Table I), where correlation coefficients are computed using Pearson’s  $r$  and it is considered as significant when its corresponding  $p$ -value  $\leq 0.05$ . Soft clothing attributes are presented into three groups, as in Table I, to highlight the mutual relations between soft attributes existing in a particular body part (i.e. *overall*, *upper*, or *lower*) and induces potential predictability of missing soft attributes. It is worth noting that

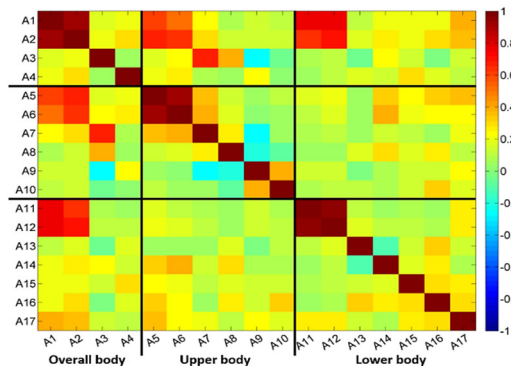


Fig. 6. Correlations between automatic soft clothing attributes.

a high positive/negative correlation between two soft attributes does not necessarily indicate that these two soft attributes are increasing or decreasing together. Rather, it indicates that they are simultaneously present in a single subject’s annotation, if one or both of the correlated pair are (non-relative) binary or multi-class soft attributes such as *Footwear category* (A17) and *Overall clothing season* (A2). Moreover, a low correlation between two soft attributes does not suggest that there is no relationship between them, but conveys a notion that this correlation is not prevalent within the clothing dataset currently used.

Within all the three body parts – *overall*, *upper*, or *lower* – the relations between skin exposure soft attributes and the clothing season represent the highest correlations, such that (A1) with (A2), (A5) with (A6), and (A11) with (A12) are strongly correlated, as expected. Besides, in the *overall* body the *Skin exposure* (A1) and *Clothing season* (A2) are well correlated with their counterparts in the *upper* body (A5 and A6) and in the *lower* body (A11 and A12). (A5) is correlated with *Lower pattern* (A16) and *Footwear category* (A17). *Upper vs Lower contrast* (A3) shows a significant correlation with *Upper brightness* (A7), and also correlates with *Upper color-scheme* (A8). This is consistent with the reality in that the *upper* (top) clothes worn by people have much more variety of colors and hues than the *lower* clothes, which in turn affect the *overall* appearance and style much more. Fig. 6 shows also that there are some soft attributes have multiple modest correlations with others like (A8, A14, A16, and A17).

#### C. Analysis of Variance (ANOVA)

One-way ANOVA is used to determine the most effective and significant automatic soft clothing traits. The F-ratio and  $p$ -value are computed per soft attribute for categorical and comparative traits, where the degree of freedom ( $df$ ) = 255 for F-ratio, and the  $p$ -value significance level is  $p \leq 0.05$ . We use the F-ratio and  $P$ -value to sort the clothing traits by their estimated capability of distinguishing between subjects, where a larger F-ratio with a smaller  $p$ -value performs better, reflecting a more successful trait in terms of discrimination between subjects. Table III reports the resulting ANOVA test values for all clothing traits ordered by the best  $p$ -value. Accordingly, Fig. 7 depicts ordered  $p$ -values scaled positively by computing the absolute logarithm of the  $p$ -value, which

TABLE III  
AUTOMATIC SOFT CLOTHING TRAITS ORDERED BY P-VALUE

(a) Categorical traits			(b) Comparative traits		
Clothing attribute	F-ratio $df = 255$	$p$ -value $p \leq 0.05$	Clothing attribute	F-ratio $df = 255$	$p$ -value $p \leq 0.05$
A12.Lower Clothing season	214.69	1.36E-111	A7	86.23	5.32E-87
A11.Lower Skin exposure	131.49	2.47E-98	A11	61.46	5.58E-78
A7.Upper Brightness	90.92	2.03E-88	A13	39.78	1.52E-66
A13.Lower Brightness	42.56	2.66E-68	A3	25.17	8.98E-55
A16.Lower Pattern	40.50	5.19E-67	A16	15.45	1.09E-42
A3.Upper vs Lower Contrast	33.59	3.73E-62	A1	11.73	3.60E-36
A14.Lower Color-scheme	28.55	5.57E-58	A10	7.25	1.56E-25
A8.Upper Color-scheme	15.18	2.96E-42	A5	3.69	6.07E-13
A10.Upper Pattern	9.03	2.97E-30	A9	3.21	8.57E-11
A4.Overall Color-scheme	8.90	6.33E-30	A15	0.82	0.86
A1.Overall Skin exposure	6.74	5.00E-24			
A5.Upper Skin exposure	4.47	3.78E-16			
A2.Overall Clothing season	4.05	1.77E-14			
A9.Upper Dominant color	3.21	8.58E-11			
A6.Upper Clothing season	3.12	2.33E-10			
A15.Lower Dominant color	0.82	8.62E-01			
A17.Footwear Category	n/a	n/a			

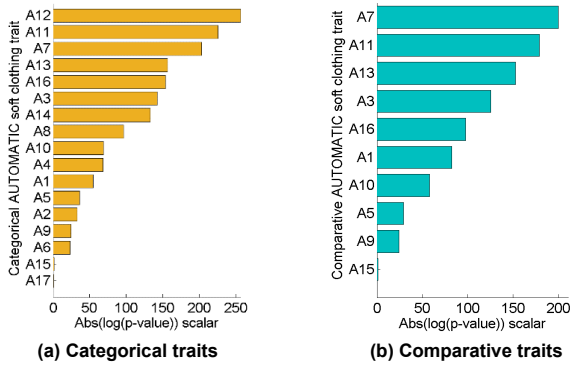


Fig. 7. Automatic soft clothing traits ordered according to p-values.

emphasizes smaller  $p$ -values by larger bars. Scaling is used to magnify small differences between  $p$ -values and to be visually observable. Among the categorical traits, *Lower clothing season* (A12) is the most discriminative trait, offering a high variance within different subjects in the database against very low variance within different (front-/side-view) samples of the same subject. *Lower skin exposure* (A11) and *Upper Brightness* (A7) are the next highly discriminative traits in their categorical and comparative forms. The traits of the *Lower brightness* (A13), *Lower pattern* (A16), and *Upper vs lower contrast* (A3) appear to be far better discriminative trait in comparison with the later traits in both the categorical and comparative forms, but in a different order. Finding (A3 and A7) within the top discriminative clothing traits, emphasizes their efficacy since they are also observed as being very well correlated and significant, as reported in Section V-B.

## VI. SOFT CLOTHING BIOMETRIC RETRIEVAL

We conduct a number of experiments to achieve soft biometric-based retrieval and we adopt a challenging methodology [19] evaluating the use of soft clothing biometrics for front/side viewpoint invariant retrieval. For all soft biometric galleries described in Table IV, all feature vectors derived by describing front-view subject images are retained in galleries to be used as enrolled biometric signatures for identity matching and retrieval. Furthermore, all feature vectors derived by describing side-view subject images are excluded from galleries and arranged into a number of *query* subsets equal to the number of galleries. Each query subset is designated to probe a corresponding targeted gallery; matching the same number, order, and types of values of the feature vector in the gallery. Subject retrieval can be distinguished from ordinary identification by the fact that it is a task that aims to identify an *unknown* subject using their biometric signature to find a match in those signatures enrolled in a database [23]. Therefore, such a retrieval task concerns the ability to generalize to *unseen* data. Viewpoint invariant subject retrieval is achieved here using the constructed query subsets, as totally new (unseen) data that is obtained from subjects' images captured from different viewpoint (i.e. side-view). As such, a query description of an (unknown) subject is used to probe a gallery by comparing and matching their biometric signature with all the signatures enrolled in the gallery. When using side-view clothing

TABLE IV  
BODY AND CLOTHING GALLERIES FOR RETRIEVAL

Clothing-based soft biometrics	
<i>Cat-6</i>	6 manual categorical clothing traits; the best correlated and most discriminative via ANOVA
<i>Cmp</i>	7 manual comparative soft clothing traits
<i>AutoCat-14</i>	Top 14 automatic categorical clothing traits via ANOVA
<i>AutoCmp-5</i>	Top 5 automatic comparative clothing traits via ANOVA
<i>Auto(Cat-14&amp;Cmp-5)</i>	Fusion of <i>AutoCat-14</i> and <i>AutoCmp-5</i>
Body-based soft biometrics	
<i>tradSoft</i>	4 categorical soft body biometrics ( <i>Age, Ethnicity, Sex, and Skin Color</i> )
<i>softBody</i>	17 categorical soft body biometrics including <i>tradSoft</i>
Combined soft clothing & body biometrics	
<i>tradAutoCat-14</i>	<i>AutoCat-14</i> combined with <i>tradSoft</i>
<i>tradAutoCmp-5</i>	<i>AutoCmp-5</i> combined with <i>tradSoft</i>
<i>tradAuto(Cat-14&amp;Cmp-5)</i>	<i>Auto(Cat-14&amp;Cmp-5)</i> combined with <i>tradSoft</i>
<i>tradCat-6</i>	<i>Cat-6</i> combined with <i>tradSoft</i>
<i>tradCmp</i>	<i>Cmp</i> combined with <i>tradSoft</i>
<i>softAutoCat-14</i>	<i>AutoCat-14</i> combined with <i>softBody</i>
<i>softAutoCmp-5</i>	<i>AutoCmp-5</i> combined with <i>softBody</i>
<i>softAuto(Cat-14&amp;Cmp-5)</i>	<i>Auto(Cat-14&amp;Cmp-5)</i> combined with <i>softBody</i>
<i>softCat-6</i>	<i>Cat-6</i> combined with <i>softBody</i>
<i>softCmp</i>	<i>Cmp</i> combined with <i>softBody</i>

descriptions to retrieve the right person enrolled by their front-view description, this raises additional challenges, emulating some real-life surveillance cases in their conditions and complexity, where a number of soft clothing attributes are likely to be occluded or difficult to observe and extract. As such, soft clothing biometrics are examined and evaluated in subject retrieval in three different aspects:

- supplementing traditional soft biometrics to enhance retrieval by soft clothing traits added to only four global soft traits (*Age, Ethnicity, Sex, and Skin Color*);
- supplementing body soft biometrics to achieve enhanced retrieval by soft clothing traits added to a subset of 17 invariant body soft traits ; and
- achieving successful retrieval by soft clothing traits alone.

Soft body descriptions are available for the current Soton dataset, which were previously collected [24], such that each of 128 subjects was labeled by multiple users describing 23 soft body traits. This body description data is used to construct two soft body biometric galleries. The first soft body gallery (*tradSoft*) comprises a four-trait feature vector for each subject; each feature vector is formed by four normalized average-labels representing the traditional biometrics (i.e. *Age, Ethnicity, Sex, and Skin Color*). The second soft body gallery (*softBody*) consists of a 17-trait feature vector per subject; each feature vector is formed using 17 soft body traits, observable from front and side viewpoints, including the former four traditional traits, where each trait is represented by a normalized average-label. Here, we utilize the performance of subject retrieval using *tradSoft* or *softBody* alone as a baseline to compare it with the performance when supplementing them with our soft clothing biometrics. In addition to the four clothing galleries (*AutoCat-14, Cat-6, AutoCmp-5* and *Cmp*), we compose the fifth hybrid gallery *Auto(Cat-14&Cmp-5)* by fusing *AutoCat-14* and *AutoCmp-5*, utilizing the potency of the high-performance traits in both categorical and comparative forms. All five soft clothing galleries are used along with the two soft body galleries (*tradSoft, and softBody*) to compose further ten combined galleries enabling to examine and evaluate the capability and performance of soft clothing biometrics in subject retrieval.

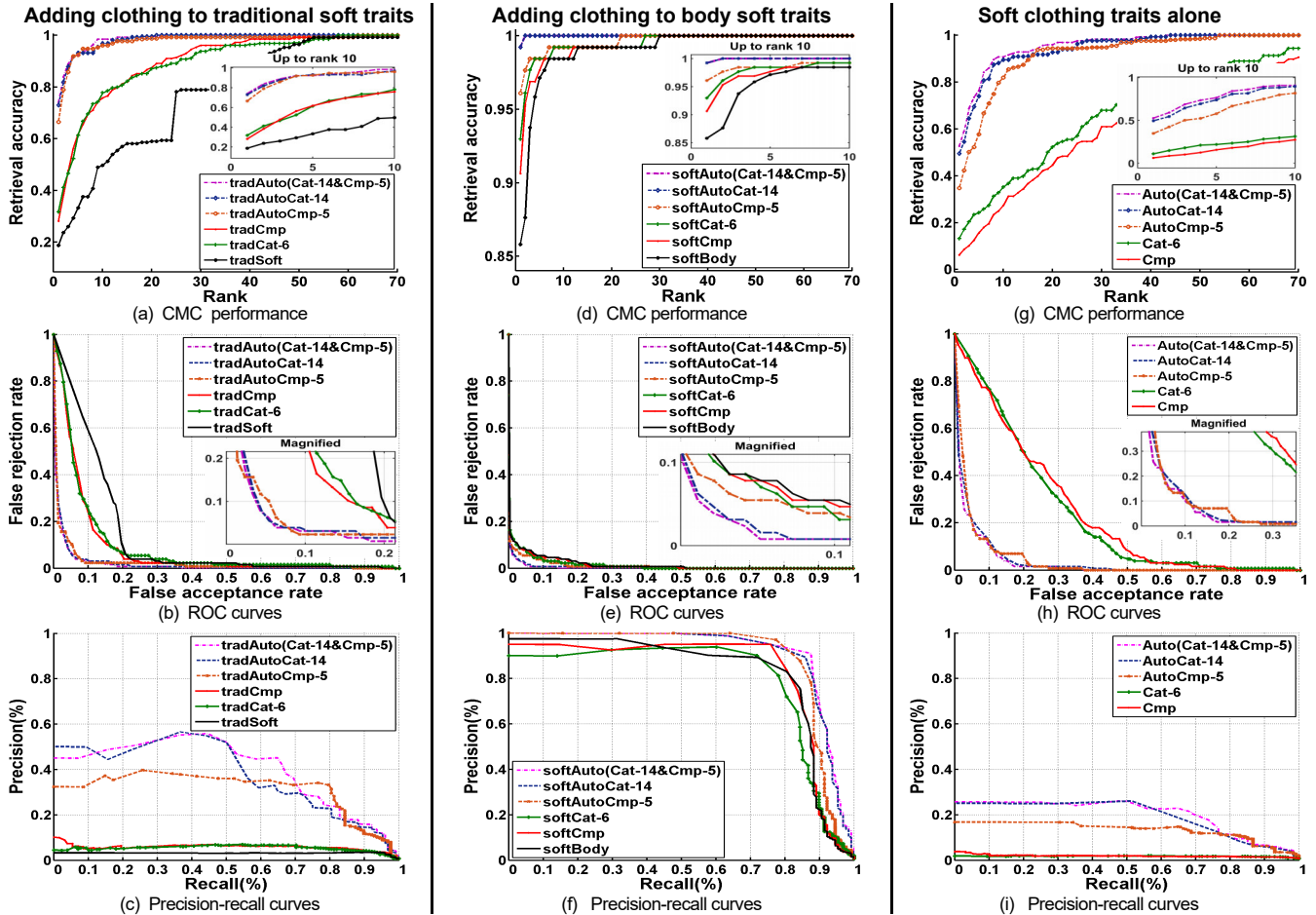


Fig. 8. CMC and ROC of retrieval performance using soft clothing biometrics (as in Table IV).

Table IV describes the ensuing biometric galleries of soft clothing-based biometrics, soft body-based biometrics, and combined soft body and clothing biometrics. The combination of two galleries is performed by consistently concatenating every single feature vector from the first gallery with a corresponding feature vector from the second gallery, in which both feature vectors belong to the same subject. For body-based galleries, a subset of subjects’ annotations is excluded from the obtained soft body data to be used as query-vectors. However, for each combined gallery, corresponding query-vectors are normalized and then reshaped and concatenated according to the feature vectors in the tested gallery to enable matching and retrieval. Based on the  $k$  nearest neighbor concept, the likelihood is estimated by the sum of Euclidean distance between each query-vector and all subject-vectors in a tested gallery, resulting in an ordered list of all subjects based on likelihood. Retrieval performance is evaluated and compared using several standard metrics: Cumulative Match Characteristic (CMC); Receiver Operator Characteristic

(ROC) with its relevant measurements of the Area Under the Curve (AUC), the Equal Error Rate (EER); the Decidability Index ( $d'$ ); and the F-measure known as (F1 score) deduced from associated *precision* and *recall* metrics. With respect to all metrics, all approaches are ranked according to their overall performance.

A. Adding Clothing to Traditional Soft Biometrics

Soft clothing biometrics are used to supplement the traditional soft biometrics (*Age, Ethnicity, Sex, and Skin Color*) to achieve enhanced person retrieval, since these four traits are commonly used in the literature and considered as essential (global) soft traits to be often embedded with other body or face soft traits, so here we embed them with clothing soft traits. Retrieval performance scores are reported in Table V for the use of traditional soft traits in isolation (i.e. *tradSoft*) in comparison with the five supplemented by clothing approaches (described in Table IV). Fig. 8-a, -b, and -c show performance curves in terms of CMC (8-a), ROC (8-b), and Precision-recall (8-c). In overview, all automatic and manual clothing approaches significantly improve the performance and offer enhancement in retrieval scores over *tradSoft*, ranging from 10% to 55% at rank 1. Generally, the automatic approaches surpass the manual approaches, as can be observed in the related graphs in Fig. 8-a, -b, and -c, and as consistently suggested by the overall rank in Table V. The *tradAuto(Cat-14&Cmp-5)* approach outperforms all its counterparts in all

TABLE V  
ADDING SOFT CLOTHING TO TRADITIONAL SOFT BIOMETRICS

Approach	Top rank	Avg sum scores up to rank		100% achieved at rank	EER	AUC	$d'$	Avg. F1	overall rank
	=1	=10	=128						
<i>tradSoft</i>	0.187	0.345	0.874	73	0.203	0.129	1.857	0.331	6
<i>tradCat-6</i>	0.318	0.595	0.940	58	0.145	0.089	1.249	0.421	5
<i>tradCmp</i>	0.281	0.588	0.944	69	0.130	0.081	1.975	0.427	4
<i>tradAutoCat-14</i>	0.729	0.899	0.991	21	<b>0.052</b>	0.019	3.086	0.595	2
<i>tradAutoCmp-5</i>	0.665	0.891	0.987	62	0.062	0.025	2.883	0.596	3
<i>tradAuto(Cat-14&amp;Cmp-5)</i>	<b>0.741</b>	<b>0.913</b>	<b>0.992</b>	23	<b>0.052</b>	<b>0.016</b>	<b>3.156</b>	<b>0.607</b>	<b>1</b>

terms but in achieving 100%, as *tradAutoCat-14* is the fastest to achieve 100% accuracy at rank 21. It also shares the same EER=0.052 with *tradAuto(Cat-14&Cmp-5)*. The ROC (Fig. 8-b) and Precision-recall (Fig. 8-c) curves show that, although *tradCat-6* and *tradCmp* have close trends in retrieval performance, *tradCmp* commits fewer ROC errors with a slightly higher overall accuracy than *tradCat-6*.

*B. Adding Soft Clothing to Body Soft Biometrics*

Soft body biometrics are supplemented using soft clothing biometrics for enhanced person retrieval, targeting the highest possible performance via the full power of soft body traits and the efficacy of soft clothing traits. CMC, ROC, and Precision-recall curves are provided in Fig. 8-d, -e, and -f respectively, demonstrating the retrieval performance of soft body traits *softBody* in isolation and when adding clothing to them through the five approaches (described in Table IV). Moreover, the consequent CMC scores and the ROC analysis of all approaches are illustrated in Table VI. By all measures, *softAuto(Cat-14&Cmp-5)* attains the highest retrieval performance receiving the best scores in all evaluation measurements. *softAutoCat-14* is the next highest performance approach with very similar scores. The accuracy rate of *softBody* jumps from 85% to 99% and rapidly reaches 100% at rank 2, when supplemented with automatic clothing traits in both approaches, *softAuto(Cat-14&Cmp-5)* and *softAutoCat-14*. The good results gained by with *softAutoCmp-5* reveal that the latent capability of comparative clothing traits in augmenting recognition or retrieval by using a minimal number of traits could be reduced to only five discriminative comparative traits as in the current case. This capability can also be inferred from the increased performance of *softAuto(Cat-14&Cmp-5)*, which results from fusing the comparative traits of *AutoCmp-5* with the categorical traits of *AutoCat-14*. Unlike when adding clothing to traditional soft traits, *softCat-6* gains a higher performance and better results than *softCmp* with respect to all evaluation measures in Table VI and performance curves in Fig. 8-d, -e, and -f.

*C. Retrieval using Soft Clothing Biometrics Alone*

The clothing-based soft traits are used in isolation, aiming to achieve successful retrieval and to investigate the max potency of pure soft clothing traits when used alone in person retrieval, which is deemed as a mimic of challenging real-case forensic scenarios when soft clothing attributes are the only observable soft traits. The retrieval evaluation results and performance analysis for all five clothing-based approaches (described in Table IV), are illustrated in Table VII along with Fig. 8-g, -h, and -i. Despite the modest retrieval rate of all approaches at rank 1 that do not exceed 53%, the retrieval rate increases sharply to score more than 77% on average at up to rank 10, especially for the three automatic approaches. Furthermore, the retrieval rate continues to increase yielding a high average match score of around 97% at the full rank 128. As when supplementing traditional and body biometrics, the *Auto(Cat-14&Cmp-5)* approach aptly retains its superiority over the other approaches and attains the best retrieval

TABLE VI  
ADDING SOFT CLOTHING TO BODY SOFT BIOMETRICS

Approach	Top rank	Avg sum scores up to rank			100% achieved at rank	EER	AUC	d'	Avg. F1	overall rank
	=1	=10	=128	=128						
<i>softBody</i>	0.858	0.952	0.995	30	0.068	0.016	3.372	0.638	6	
<i>softCat-6</i>	0.930	0.978	0.997	27	0.063	0.013	3.082	0.639	4	
<i>softCmp</i>	0.906	0.970	0.996	30	0.067	0.015	3.343	0.638	5	
<i>softAutoCat-14</i>	<b>0.992</b>	<b>0.999</b>	<b>0.999</b>	<b>2</b>	<b>0.032</b>	0.004	3.914	0.660	2	
<i>softAutoCmp-5</i>	0.961	0.985	0.998	22	0.054	0.010	3.663	0.650	3	
<i>softAuto(Cat-14&amp;Cmp-5)</i>	<b>0.992</b>	<b>0.999</b>	<b>0.999</b>	<b>2</b>	<b>0.032</b>	<b>0.003</b>	<b>4.009</b>	<b>0.663</b>	<b>1</b>	

TABLE VII  
USING SOFT CLOTHING BIOMETRICS ALONE

Approach	Top rank	Avg sum scores up to rank			100% achieved at rank	EER	AUC	d'	Avg. F1	overall rank
	=1	=10	=128	=128						
<i>Cat-6</i>	0.133	0.224	0.806	100	0.303	0.232	0.630	0.250	4	
<i>Cmp</i>	0.063	0.166	0.773	105	0.319	0.240	1.056	0.275	5	
<i>AutoCat-14</i>	0.495	0.740	0.968	44	0.102	0.039	2.496	0.414	2	
<i>AutoCmp-5</i>	0.347	0.612	0.954	54	0.100	0.043	2.409	0.519	3	
<i>Auto(Cat-14&amp;Cmp-5)</i>	<b>0.527</b>	<b>0.772</b>	<b>0.975</b>	<b>43</b>	<b>0.099</b>	<b>0.033</b>	<b>2.581</b>	<b>0.547</b>	<b>1</b>	

performance. This is reflected by all associated curves in Fig. 8 and metrics in Table VII. Though *AutoCat-14* is second according to the overall rank, *AutoCmp-5* receives better scores in terms of EER, indicating smaller errors, as can also be observed in the magnified box of the ROC curves in Fig. 8-h. It also achieves a better average F1 score, considering the trade-off between the estimated precision and recall represented in Fig. 8-i. *Cat-6* also surpasses *Cmp* here in all CMC and ROC metrics, but *d'* and F1, which are observed to be better and higher by *Cmp*. In all aspects when soft clothing are used, the average F-measure receives low scores compared with CMC since there is likely to be a large number of false positives against only few false negatives.



Fig. 9. Examples of applying our automatic clothing attribute annotation on images (captured in the wild) from the Clothing Attribute Dataset [1].

VII. CONCLUSIONS

We show that it is possible to use labels and automatically derived clothing descriptions for human recognition purposes. This is a precursor to automated search by clothing for forensic use. Our proposed approach enables subject identification and retrieval by using either an image or a verbal description, which reflects forensic scenarios where a suspect's image or an eyewitness statement is available. Soft clothing traits could be very useful as a major cue or ancillary information for identity in scenarios suffering from high variability issues. Since clothing information are more likely visible and perceivable soft biometrics in surveillance videos, they could be effective in challenging cases such as when criminals cannot be identified by their faces.

The capability of retrieving a subject-of-interest within a relatively small list (e.g. 10 subjects) indicates a promising potential of effective use in various biometric applications.

This can be beneficial in narrowing the search to a much smaller number of candidates. Enhanced retrieval performance attained by fusing the best categorical and comparative clothing traits, referred to as *Auto(Cat-14&Cmp-5)*, signifies that categorical and comparative traits are two different measures. Although they both describe the same relative soft attribute, they offer different information on identity in a way that is mutually supplemental. Relative soft clothing traits appear to be more significant in composing a discriminative biometric signature of the subject.

Naturally, we intend in future to translate this to images derived in the wild. Fig. 9 shows the result of applying the automatic clothing attribute annotation to images of subjects in the Clothing Attribute Dataset. Here we show that it is possible to handle complex backgrounds and lighting so as to derive the clothing annotations for use in identification from such images. This paper describes the validity of the approach, in that it is indeed possible to use clothing annotations and attributes for identification purposes.

#### REFERENCES

- [1] H. Chen, A. Gallagher, and B. Girod, "Describing clothing by semantic attributes," in *ECCV*, 2012, pp. 609-623.
- [2] K. Yamaguchi, T. Okatani, K. Sudo, K. Murasaki, and Y. Taniguchi, "Mix and Match: Joint Model for Clothing and Attribute Recognition," in *BMVC*, 2015.
- [3] L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. Van Gool, "Apparel classification with style," in *ACCV*, 2013.
- [4] J. Huang, R. S. Feris, Q. Chen, and S. Yan, "Cross-domain image retrieval with a dual attribute-aware ranking network," in *ICCV*, 2015.
- [5] S. Vittayakorn, K. Yamaguchi, A. C. Berg, and T. L. Berg, "Runway to realway: Visual analysis of fashion," in *WACV*, 2015.
- [6] M. Yang and K. Yu, "Real-time clothing recognition in surveillance videos," in *ICIP*, 2011, pp. 2937-2940.
- [7] S. Liu, J. Feng, Z. Song, T. Zhang, H. Lu, C. Xu, et al., "Hi, magic closet, tell me what to wear!," in *ACM MM12*, 2012.
- [8] S. Liu, Z. Song, G. Liu, C. Xu, H. Lu, and S. Yan, "Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set," in *CVPR*, 2012.
- [9] A. Dantcheva, C. Velardo, A. D'angelo, and J.-L. Dugelay, "Bag of soft biometrics for person identification," *Multimedia Tools and Applications*, vol. 51, pp. 739-777, 2011.
- [10] G. Mu, D. Yuning, A. Haizhou, and L. Shihong, "A hybrid approach to pedestrian clothing color attribute extraction," in *IAPR MVA*, 2015.
- [11] S. Denman, M. Halstead, C. Fookes, and S. Sridharan, "Searching for people using semantic soft biometric descriptions," *Pattern Recognition Letters*, vol. 68, pp. 306-315, 2015.
- [12] M. Halstead, S. Denman, S. Sridharan, and C. Fookes, "Locating people in video from semantic descriptions: A new database and approach," in *ICPR*, 2014.
- [13] Z. Shi, T. M. Hospedales, and T. Xiang, "Transferring a semantic representation for person re-identification and search," in *CVPR*, 2015.
- [14] Q. Chen, J. Huang, R. Feris, L. M. Brown, J. Dong, and S. Yan, "Deep Domain Adaptation for Describing People Based on Fine-Grained Clothing Attributes," in *CVPR*, 2015.
- [15] D. A. Vaquero, R. S. Feris, D. Tran, L. Brown, A. Hampapur, and M. Turk, "Attribute-based people search in surveillance environments," in *WACV*, 2009.
- [16] A. Kovashka, D. Parikh, and K. Grauman, "Whittlesearch: Image search with relative attribute feedback," in *CVPR*, 2012, pp. 2973-2980.
- [17] E. S. Jaha and M. S. Nixon, "Soft Biometrics for Subject Identification using Clothing Attributes," in *IJCB*, 2014.
- [18] E. S. Jaha and M. S. Nixon, "Analysing Soft Clothing Biometrics for Retrieval," in *BIOMET*, 2014.
- [19] E. S. Jaha and M. S. Nixon, "Viewpoint invariant subject retrieval via soft clothing biometrics," in *ICB*, 2015, pp. 73-78.
- [20] M. Ao, D. Yi, Z. Lei, and S. Z. Li, "Face recognition at a distance: System issues," in *Handbook of Remote Biometrics*, ed. 2009.
- [21] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? A survey on soft biometrics," *IEEE TIFS*, 2015.
- [22] U. Park and A. K. Jain, "Face matching and retrieval using soft biometrics," *IEEE TIFS*, vol. 5, pp. 406-415, 2010.
- [23] D. Reid, M. Nixon, and S. Stevenage, "Soft Biometrics; Human Identification using Comparative Descriptions," *IEEE TPAMI*, vol. 36, pp. 1216-1228, 2014.
- [24] S. Samangooei and M. S. Nixon, "Performing content-based retrieval of humans using gait biometrics," *Multimedia Tools and Applications*, vol. 49, pp. 195-212, 2010.
- [25] P. Tome, J. Fierrez, R. Vera-Rodriguez, and M. Nixon, "Soft Biometrics and Their Application in Person Recognition at a Distance," *IEEE TIFS*, vol. 9, pp. 464-475, 2014.
- [26] J. Thornton, J. Baran-Gale, D. Butler, M. Chan, and H. Zwahlen, "Person attribute search for large-area video surveillance," in *HST*, 2011.
- [27] D. Adjeroh, C. Deng, M. Piccirilli, and A. Ross, "Predictability and correlation in human metrology," in *WIFS*, 2010.
- [28] K. Niinuma, U. Park, and A. K. Jain, "Soft biometric traits for continuous user authentication," *IEEE TIFS*, pp. 771-780, 2010.
- [29] M. Demirkus, D. Precup, J. J. Clark, and T. Arbel, "Soft biometric trait classification from real-world face videos conditioned on head pose estimation," in *IEEE CVPRW* 2012, pp. 130-137.
- [30] C. Deng, C. Cunjian, M. Piccirilli, D. Adjeroh, T. Bourlai, and A. Ross, "Can facial metrology predict gender?," in *IJCB*, 2011.
- [31] P. Layne, T. M. Hospedales, and S. Gong, "Person Re-identification by Attributes," in *BMVC*, 2012.
- [32] J. Zhu, S. Liao, Z. Lei, D. Yi, and S. Z. Li, "Pedestrian Attribute Classification in Surveillance: Database and Evaluation," in *ICCV*, 2013.
- [33] M. S. Nixon, P. Correia, K. Nasrollahi, T. Moeslund, A. Hadid, and M. Tistarelli, "On Soft Biometrics," *Pattern Recognition Letters*, 2015.
- [34] H. Wang, X. Bao, R. R. Choudhury, and S. Nelakuditi, "InSight: recognizing humans without face recognition," in *WMCSA*, 2013.
- [35] A. Li, L. Liu, K. Wang, S. Liu, and S. Yan, "Clothing Attributes Assisted Person Re-identification," *IEEE TCSVT*, 2014.
- [36] A. Schumann and E. Monari, "A soft-biometrics dataset for person tracking and re-identification," in *AVSS*, 2014.
- [37] X. Chen, L. An, and B. Bhanu, "Soft-Biometrics and Reference Set Integrated Model for Tracking Across Cameras," in *Distributed Embedded Smart Cameras*, ed. Springer, 2014, pp. 211-230.
- [38] S. Sunderrajan and B. Manjunath, "Context-Aware Hypergraph Modeling for Re-identification and Summarization," *IEEE Transactions on Multimedia*, vol. 18, pp. 51-63, 2016.
- [39] Y.-F. Wang, E. Y. Chang, and K. P. Cheng, "A video analysis framework for soft biometry security surveillance," in *VSSN*, 2005.
- [40] A. C. Gallagher and T. Chen, "Clothing cosegmentation for recognizing people," in *CVPR*, 2008.
- [41] S. Cai, J. Wang, and L. Quan, "How Fashion Talks: Clothing-Region-Based Gender Recognition," *CIARP*, pp. 515-523, 2014.
- [42] M. Toews and T. Arbel, "Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion," *IEEE TPAMI*, vol. 31, pp. 1567-1581, 2009.
- [43] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *PETS*, 2007.
- [44] M. D. MacLeod, J. N. Frowley, and J. W. Shepherd, "Whole body information: its relevance to eyewitnesses," in *Adult eyewitness testimony*, ed. Cambridge Uni. Press, 1994.
- [45] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE TPAMI*, vol. 32, pp. 1627-1645, 2010.
- [46] P. Jones, P. Viola, and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *CVPR*, 2001.
- [47] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, "Progressive search space reduction for human pose estimation," in *CVPR*, 2008.
- [48] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *ACM TOG*, vol. 23, pp. 309-314, 2004.
- [49] D. A. Winter, *The Biomechanics and Motor Control of Human Movement*, 2nd ed.: John Wiley & Sons, 1990.
- [50] C. O. Conaire, N. E. O'Connor, and A. F. Smeaton, "Detector adaptation by maximising agreement between independent data sources," in *CVPR*, 2007.
- [51] D. E. Ilea and P. F. Whelan, "Color image segmentation using a spatial k-means clustering algorithm," in *IMVIP*, 2006.
- [52] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE TPAMI*, vol. 24, pp. 971-987, 2002.
- [53] T. Joachims, "Optimizing search engines using clickthrough data," in *ACM SIGKDD*, 2002, pp. 133-142.
- [54] D. Parikh and K. Grauman, "Relative attributes," in *ICCV*, 2011.
- [55] J. Shutler, M. Grant, M. S. Nixon, and J. N. Carter, "On a large sequence-based human gait database," in *RASC*, 2002.