UNIVERSITY OF SOUTHAMPTON

Soft Biometrics using Clothing Attributes for Human Identification

by

Emad Sami Jaha

A thesis submitted for the degree of Doctor of Philosophy

in the

Faculty of Physical and Applied Sciences

School of Electronics and Computer Science

May 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL AND APPLIED SCIENCES
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy
by Emad Sami Jaha

Recently, soft biometrics has emerged as a novel attribute-based person description for identification. It is likely that soft biometrics can be deployed where other biometrics cannot, and have stronger invariance properties than traditional vision-based biometrics, such as invariance to illumination and contrast. Previously, a variety of soft body and face biometrics have been used for identifying people and have increasingly garnered more research interest and are often considered as major cues for identity, especially in the absence of valid traditional hard biometrics, as in surveillance.

Describing a person by their clothing properties is a natural task performed by people. As yet, clothing descriptions have attracted little attention for biometric purposes as it has been considered unlikely to be a potential cue to identity. There has been some usage of clothing attributes to augment biometric description, but a detailed description has yet to be used. In everyday life, several cases and incidents arise highlighting the usefulness and capability of information deduced from clothing regarding identity. Clothing is inherently more effective for short-term identification, since people can change clothes.

This thesis introduces semantic clothing attributes as a new form of soft biometrics. The usability and efficacy of a novel set of proposed soft clothing traits is explored, showing how they can be exploited for human identification and re-identification purposes. Furthermore, the viability of these traits is investigated in correctly retrieving a subject of interest, given a verbal description of their clothing. The capability of clothing information is further examined in more realistic scenarios offering viewpoint invariant subject retrieval.

Although clothing traits can be naturally described or 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 thesis proposes further a novel set of automatic 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 clothing attributes inferred by human vis-a-vis those which are inferred automatically by computer-vision.

Extended analysis of clothing information is conducted. Human identification and retrieval are achieved, evaluated, and compared using different proposed forms of soft clothing biometrics in addition and in isolation. The experimental results of identification and retrieval highlight clothing attributes as a potentially valuable addition to the field of soft biometrics.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ANOVA	Analysis of Variance
AUC	Area under the Curve
CAP	Clothing Annotation Procedure
CCTV	Closed Circuit Television
CMC	Cumulative Match Characteristic
EER	Equal Error Rate
F1	F-measure deduced from precision and recall metrics
FAR	False Acceptance Rate
FRR	False Rejection Rate
Grabcut	Grab and cut technique for foreground object extraction
Kappa	Cohen's kappa statistical consensus measurement
k-means	k-means clustering method of vector quantization
***k*NN**	*k* Nearest Neighbour classifier
LBP	Local Binary Pattern
MANOVA	Multivariate analysis of variance
MI	Mutual Information
NLP	Natural Language Processing
Pearson's *r*	Pearson's product-moment correlation coefficient
Prewitt	Prewitt operator used within edge detection algorithms
ROC	Receiver Operator Characteristic
SD	Statistical Dependency
SFFS	Sequential Floating Forward Selection
SFS	Sequential Forward Selection
SGDB	Soton Gait Database
SVM	Support Vector Machines
ULBP	Uniform Local Binary Pattern

DECLARATION OF AUTHORSHIP

I, Emad Sami Jaha declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University;
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- Where I have consulted the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Singed: ______________________________

Date: ______________________________

ACKNOWLEDGMENTS

First and foremost, all my thanks are due to Almighty Allah, the most merciful and gracious, for countless blessings and reasons, which are endless to mention.

I would like to express my sincere thanks and gratitude to my supervisor Professor Mark S. Nixon for giving me the opportunity to carry out this research and explore a novel idea in an interesting field. His precious time, wide experience, insightful guidance, and continuing support all made this research possible and meaningful. I indeed had the privilege of being one of his supervisees.

I would like to express my countless thanks to my honourable parents, as no words can describe my gratitude to them, for all what they did for me. Their love, support, sacrifice, and care for me when I was little are what made all this possible.

I would like to deeply thank my dear wife and share her any success, as she was always my sincere companion sharing with me all moments, feelings, and challenges during my PhD journey. I would like to extend my thanks to my beloved children for their patience along with my apologies for being so busy during the last few years and not giving them the time they deserve.

My thanks also go to all my brothers and sisters who were always encouraging and supportive of me during my academic life. Besides, my colleagues and friends inside and outside the university who shared with me the path of learning and knowledge throughout my postgraduate studies.

Finally, my especial thanks are to King Abdulaziz University for their generous support and scholarship to pursue my PhD study and to the University of Southampton for their facilities and support of this research.

Chapter 1

Introduction

1.1 Context and Contributions

1.1.1 Motivations and Applications

In everyday life, people use clothing descriptions to identify/re-identify each other, especially from a distance or when the faces are not visible. Furthermore, people can be unique by clothing and many people often wear similar clothing or a certain clothing style. 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]. This can be exploited when composing a biometric signature and even further to possibly infer some unknown attributes from the known ones.

Figure 1.1: An image highlighting a marked suspect with covered face and distinct clothing [1], and bottom right a face image of a suspect appearing to wear the same clothes [2]

Fig. 1.1 shows an image of rioters in London 2011 and it highlights a suspect with covered face and head. No soft traits are observable except clothing attributes. Also in the bottom right corner, a CCTV image released later within a list of most wanted suspects shows

[1] The Advertiser: http://www.adelaidenow.com.au/news/world/profiles-of-london-locations-where-riots-have-broken-out/story-e6frea8l-1226111443592

[2] BBC News: http://www.bbc.co.uk/news/uk-england-london-16171972

clearly what appears to be the same clothing, suggesting an identification link to the rioter. Such an image provides a real example of how clothing attributes could be beneficial in identification and also demonstrates that, in some cases, clothing attributes can be the only observable soft traits to be exploited. In the worst case scenario, clothing can be used to narrow the search.

Several incidents and forensic scenarios 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. 1.2 the police publicized surveillance footage of the incident via social media asking whether anyone could recognize these suspects by their clothing, bags, or behaviour. As such real-life examples suggest that soft clothing attributes are likely to be useful, viable, and effective traits for characterizing an individual.

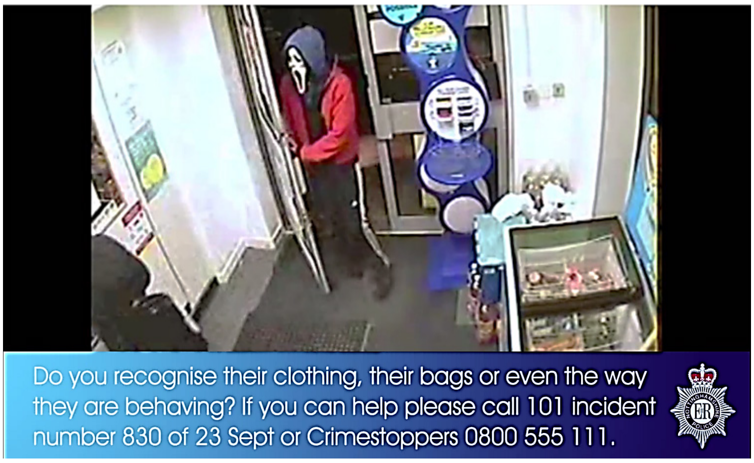


Figure 1.2: CCTV shows masked raiders (Nottingham 2015), released by police asking help to recognize them by clothing, bags, or behaviours [3]

The use of clothing is emphasized in the scope of person re-identification, which tends to be as one likely application can rely on soft clothing traits. It is rather that clothing appears

[3] Nottinghamshire Police YouTube channel: https://www.youtube.com/user/nottinghampolice

to be more effective for short-term identification/re-identification, since people change clothes. Furthermore, in case of identification or re-identification at a distance, clothing attributes are most likely usable whilst the majority of traditional attributes and biometrics are unusable. Biometric systems utilising soft clothing attributes can be widely deployed in forensic and security applications. Hence, the use of clothing for biometric purposes deserves more research interest than it has yet received.

1.1.2 From Clothing to Identity

- **Human-based (manual) soft clothing biometrics**

Although human clothes are a predominant visible characteristic of the person's appearance, they have yet to be adopted for representing soft biometric traits for an individual and have been considered unlikely as a cue to identity. Detecting the presence of some common clothing attributes, besides other soft biometrics, can supplement the low-level features used for person re-identification [3]. This allows use of more of the information available in surveillance video which is consistent with analysing data of such poor quality.

Other than biometrics, there has been previous work on recognizing clothing categories [4], semantically describing clothing pieces [1] and automatically detecting then classifying certain semantic clothing attributes in pedestrian data images [5], or alternatively classifying overall clothing styles in natural scenes based on a group of defined common categories and attributes [6]. Also automatic search and retrieval by clothing attributes for occasion-style recommendation [7] or a body attribute-based query integrated with clothing colours and types for a people search in surveillance data [8].

There is emergent work in using computer vision for automated clothing analysis of low- to medium-level features and very few high-level features (attributes). This is generally performed and could be practical on high resolution imagery with high contrast illumination and has yet to be efficiently and reliably applied to unconstrained surveillance imagery. Since surveillance imagery is often captured in low resolution and poor quality, it will be difficult to analyse clothing information, especially if they are meant to be used as cues for identity. As such, attribute based approaches utilising human-vision, appear more suited to analysis of such surveillance imagery. In such imagery, whilst obtaining identifiable faces may be impossible, clothing appearance may become the main cue to identity [9]. Even though clothing is innately more efficient in short-term id/re-id as people might change their clothes, the regions of clothing in images/videos can offer useful extra information about the identity of the individual. Moreover, for images captured on the same day, areas associated with clothing of these images may contain substantial information helping the discrimination between people, and 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 prefers a specific clothing style or colour [10].

In this research, unlike most of the current state-of-the-art research approaches, with a view to exploiting human-based clothing descriptions into a biometric signature, we propose a group of semantic clothing attributes as soft biometric traits and explore the validity and efficiency of clothing descriptions in human identification and retrieval.

- **Vision-based (automatic) soft clothing biometrics**

Learning reliable biometric traits is required for realistic scenarios, such as after a change of viewpoint and partial occlusion [11]. In such scenarios, even some soft biometrics may likely be more vulnerable, especially to annotation subjectivity and missing information mostly caused by occlusion. Viewpoint invariance is a challenging problem that has been considered in most biometric modalities. For example a recognition algorithm suited for front-to-side re-identification, utilizing colour- and texture-based soft traits extracted from patches of hair, skin and clothes [12]. Subject retrieval is deemed as viewpoint invariant, if it remains invariant to any angle from which a subject is likely to be seen [13], such as front and side views (see Fig. 1.4). Although 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.

A major distinction of this work is that even though computer-vision is employed in extracting and analysing 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, 2, 4, 6, 7, 14-17]. Therefore, those studies without biometric interest are very slightly related to this work. This research proposes high-level clothing descriptions that are closer to 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.

For example, in describing the colour attributes of a piece of clothing, some prior research used computer vision features or descriptors to categorize clothing to a limited number of basic colour names (i.e. red, blue, black, etc.) serving as categorical labels [8, 14, 15, 18-22]. We rather use computer vision features to categorize clothing colours according to a more generic and higher level annotation describing the *Colour-scheme* as 'Cool' like blue, 'Warm' like orange, and 'Neutral' like grey, so it is less error-prone in case of illumination variation. Another distinction of our computer vision approach 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 [23]. 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. Fig. 1.3 demonstrates that this research embraces clothing attributes, soft biometrics, human identification, and computer vision.

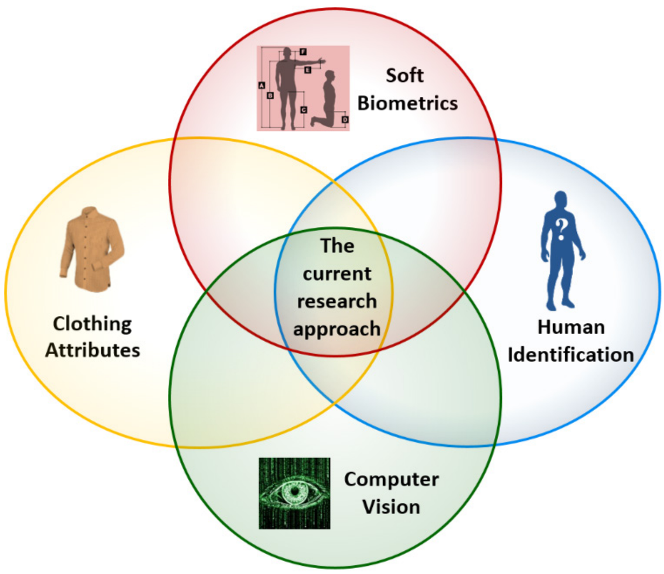


Figure 1.3: The relevant areas of the current research approach

The novelty in this research 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, as illustrated in Fig. 1.4. Our approach utilizes more fine-grained soft clothing attributes rather than coarse-grained attributes as in [21]. 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 [19]. Prior to this research, clothing has not been used purely as a biometric either by computer vision, or by semantic annotation.

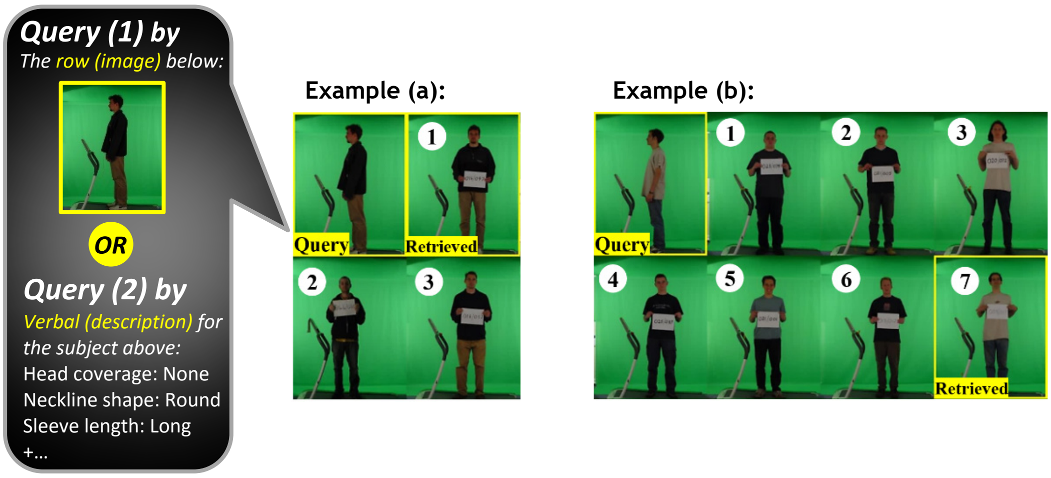


Figure 1.4: Soft biometric query for subject retrieval using either an unseen row *image* or verbal *description*, Example (a): a subject correctly retrieved at rank 1, Example (b): a subject retrieved but at rank 7

1.1.3 Major Contributions

In this research, two different groups of soft clothing biometrics are analysed and 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. Thus, a summary of the main contributions of this thesis, introducing the first approach for identifying people by their clothing as a major cue and defining this new form of cues as ***soft clothing biometrics***, this comprises:

a) **human-based (manual) soft clothing biometrics**

- new soft biometrics using semantic clothing attributes;
- increased discriminatory comparative traits derived using a Ranking SVM for clothing attributes;
- a new web-based mechanism for obtaining and analysing human-based clothing labels and comparisons;
- new biometric techniques using categorical and comparative soft clothing traits for human identification and subject retrieval in fusion and in isolation; and
- viewpoint invariant subject retrieval using soft clothing biometrics.

b) **vision-based (automatic) soft clothing biometrics**

- introducing a set of automatic soft clothing attributes to bridge the semantic gap between vision-based low-level features and high-level clothing annotations;
- a new vision-based approach for automatic soft clothing attribute annotation, 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 soft clothing biometrics in fusion and in isolation.

c) **comprehensive analysis, validation, and comparison of manual and automatic soft clothing biometrics**

- an extended analysis of the viability, correlations, and significance of soft clothing traits in identification and retrieval;
- detailed performance assessment and comparison between different forms of soft clothing traits; and
- an empirical investigation of the efficacy of soft clothing traits for feature subset selection and performance improvement.

1.1.4 Publications

Papers resulting from this work are listed chronologically below:

[P1] E. S. Jaha and M. S. Nixon, "Soft Biometrics for Subject Identification using Clothing Attributes," in *IEEE/IAPR International Joint Conference on Biometric (IJCB)*, 2014.

[P2] E. S. Jaha and M. S. Nixon, "Analysing Soft Clothing Biometrics for Retrieval," in *International Workshop on Biometrics (BIOMET)*, 2014.

[P3] E. S. Jaha and M. S. Nixon, "Viewpoint Invariant Subject Retrieval via Soft Clothing Biometrics", in *IEEE/IAPR International Conference on Biometrics (ICB)*, 2015.

[P4] E. S. Jaha and M. S. Nixon, "From Clothing to Identity: Manual and Automatic Soft Biometrics", in *IEEE Transactions on Information Forensics and Security (TIFS)*, vol. 11, no. 10, pp. 2377-2390, 2016.

[P5] E. S. Jaha and M. S. Nixon, "Clothing Analysis for Subject Identification and Retrieval", in *Springer book: Recent Advances in Intelligent Image Search and Video Retrieval*, in progress.

1.2 Human Identification

Identifying people is an important task in daily life. It is an urgent need and a routine task performed for different purposes mostly for security such as restricted access, border control and crime detection or prevention. The continuing technological advances continue to provide useful functional solutions for people identification. Many commonly-used solutions, for example username/password and smart cards, can suffer from violation and possible abuse [24].

Biometrics can solve these difficulties and provide effective automated solutions as they mostly use very discriminative traditional (hard) biometrics. Traditional biometric solutions exploit unique and inherent personal characteristics such as fingerprint, iris, and face (known as physical biometrics), or individual behaviours such as signature, voice, and gait (known as behavioural biometrics). Such classic physical and behavioural biometrics have been widely and effectively devoted to people identification and authentication.

In traditional biometric systems a person's traits need to be correctly enrolled into the database for successful use in biometric recognition or verification [25], which absolutely requires their cooperation. That is owing to the fact that such primary human traits can be influenced by many factors, such as distinctiveness, collectability, acceptability, universality, permanency and resistance to circumvention [26]. In addition, there are still several challenges and limitations to be confronted such as lower resolution and increased distance between the camera/sensor and the captured subject, where such hard traits lose utility [27]. 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. Soft biometric traits are robust against viewpoint change and appearance variation [28]. In contrast with most 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.

1.3 Soft Biometrics

Soft biometrics utilizes conventional human descriptions and translates these descriptions to the machine's biometric forms (in a way bridging the semantic gap) [29], thereby they can be used to characterize the person's face, body and accessories [30]. Soft biometric techniques mainly depend on defining a number of semantic attributes and assigning a set of descriptive labels (traits) for each attribute. Take for example, a semantic attribute "Height" which can be assigned a set of labels like ("Very short", "Short", "Average", "Tall" and "Very tall"). A semantic attribute can be any observable property that has a designated name or description by human like height, weight, gender, race, age, and eye or hair colour. Such attributes can be either binary attributes associated with categorical traits or relative attributes, associated with categorical or comparative traits. So far, the relative descriptions have been found to be more precise and informative compared with binary descriptions [31]. Apart from soft biometric use, some semantic attributes describing a , which are subjective inherently subjective

Despite of the lack of discriminatory capability associated with each single soft trait, its power can emerge when it is combined with other soft traits to be used as a biometric signature for identification [27], re-identification and database retrieval [32], or they can be used to augment other traditional physical and behavioural (hard) biometrics such as facial traits [33] or a gait signature [29, 34]. Soft traits could be automatically extracted and analysed then could be used to complement and assist some other primary traits in the recognition process [26]. Relative attributes are not only measurable representing the strength of attributes but also comparable allowing for more precise descriptions [31], which can be used to detect the small differences by describing the strength of attributes of a subject compared with others [35].

Soft body and face biometrics have been attracting increasing research interest and are often considered as major cues for identifying individuals [27, 36-40], especially in the absence of valid traditional hard biometrics. The basic approach uses human vision wherein labellers describe human body features using human understandable labels and measurements, which in turn allow for recognition and retrieval using only verbal descriptions as the sole query [41, 42]. The features also allow prediction of other measurements as they have been observed to be correlated [43]. 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 [39], it could be extended, in a more challenging application, for retrieval from video footage [27]. 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. It is desirable for a multi-attribute retrieval approach to consider correlations and interdependencies among the query terms leading to refined search and enhanced retrieval results [44].

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 [29] or soft face profile [45] with gait signature, and using soft facial traits along with other (hard) facial traits [33]. Nevertheless, soft comparative labels have been demonstrated to be more successful in representing the slight differences between people in body descriptions [41] or can be also applicable on facial descriptions [46].

In contrast to traditional face biometrics, facial soft biometrics are collectable in the absence of consent and cooperation of the surveillance subject, allowing for fast and enrolment-free biometric analysis [47]. 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 [38, 48]. Measured facial information might be useful for gender prediction [49] and many system issues and challenges could arise when soft facial traits are used at a distance [36]. Face aging is a major challenge for any face recognition technique, and it becomes more challenging when added to unconstrained imaging (in the wild) even if it involves human intervention and decision [50]. Therefore, soft facial attributes and features need to be carefully defined to be as much as possible describable and resistant against aging and other confounding factors, as such they can be more reliable and helpful for the purpose of subject identification in criminal investigations [51]. Complementing traditional face recognition systems by soft face biometrics is deemed to be a major research direction of several recent techniques [52] and a promising method significantly reducing recognition errors [53] and increasing the performance even under many complex conditions with a larger degree of variability in lighting, pose, expression, occlusion, face size and distance from the camera [54].

In hybrid biometric systems based on multiple primary traits such as faces and fingerprints, a number of soft traits can complement the identity information provided by the primary traits to improve the recognition performance [55]. For surveillance purposes, different forms of soft biometrics take place in a variety of applications and scenarios [8, 19, 20, 27, 28, 40, 56-58], where many of those soft traits could be easily distinguished even at a distance and then to be fused with vision features for target tracking [28] or to be fused with classic biometric traits, for the sake of improvement in overall recognition accuracy, especially with poor quality surveillance videos [40]. In surveillance images and videos, people's clothes are considered good discriminative features amongst the context-based information for distinguishing among people [59] and assisting identification [60]. A greater challenge for person re-identification exists across multi-camera surveillance, where typically assuming that individuals wear the same clothing between sightings, and likely some clothing attributes like colour and texture are observable with other soft body traits like height [61]. Note that age and sex are considered as important soft biometric traits for video surveillance [62]. Numerous recent approaches for people re-identification have been reviewed, discussed, and compared [63].

1.4 Clothing and Semantic Attributes

Clothing cannot just be considered to be individual, but can also reflect some cues regarding social status, lifestyle and cultural affiliation [6]. Clothing encodes more information about an individual, beyond just their visual appearance [64]. It is common for sections of societies to wear similar types of clothing, and some clothing styles appear to be correlated with age [10]. Youth cultures in particular are known to favour a particular sense of dress, such as punk, mod and goth, and there are many others. Clothing features may provide, in some sense, a number of indicators about potential behaviours or trends of a person. From the psychological perspective, someone's behaviour can be correlated with some clothing features and can be affected by certain clothing descriptions such as dark or black colour, as reported in some studies that wearing dark clothing could lead to more aggressiveness [64].

1.4.1 Clothing Analysis for Different Use

Apart from biometric use, clothing details and features have been mostly used in automated search procedures, based on computer vision modelling or machine learning techniques, through a variety of processes including detection, segmentation, feature extraction, categorization, clustering and retrieval [1, 4, 6, 7, 14, 65-67]. With the continuing increase in research interest in clothing, a number of clothing datasets, some of which are publically available, have been introduced and used mostly for vision based clothing detection and recognition for purposes like classification in natural scenes [6] or search matching and retrieval [1, 4, 7, 14]; other approaches have been reviewed in [68]. Interestingly, a keen interest in semantic clothing description is reflected in the new and public Clothing Attribute Dataset [1]. APiS is a non-public database which has been designed to evaluate a classification method for pedestrian multi-class attributes including few clothing descriptions such as upper/lower clothing categories and colours [5]. Some other pedestrian datasets like i-LIDS and VIPeR have a variety of clothing representations that have been utilized to extract some clothing attributes for different analysis and purposes.

People choose and buy clothing via semantic attributes. Defining and utilizing a list of clothing attributes for various purposes has been the concern of several non-biometric studies [1, 2, 4, 6, 7, 15-17]. Visual and semantic clothing attributes can be used and applied in real-time frameworks with different aims including: semantic clothing segmentation [66]; clothing fashion style recommendation via given user-input occasion [7]; or by captured user-image used for matching and comparisons [65, 67]; and online fashion search and retrieval [4]. Moreover, other than people recognition or tag-and-track analysis, in surveillance videos, real-time visual clothing segmentation and categorization can be performed, taking advantage of automatic face detection to specify where clothing is likely to be located [17].

1.4.2 Clothing Analysis for Identity

There are few research studies associated with using clothing for biometric purposes [8, 9, 12, 33, 42, 58, 69-74] and just few of studies associated with adopting and devoting a list of clothing attributes [3, 5], while a number of most recent relevant and partly related approaches were reviewed in [75] and [76] respectively. 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 [33, 74]; soft attributes for re-identification [3, 21] along with person detection [19, 20] and tracking [71, 77] or attribute-based people search [19, 21]; detecting and analysing semantic descriptions (labels) of clothing colours and types to supplement other body and facial soft attributes in automatic search and retrieval [8] or in automated person categorization [69]; utilising some clothing attributes like colour [58, 70] and style to improve the observation and retrieval at a distance in surveillance environments [42]; exploiting clothing colours and decorations to supplement other behavioural biometrics like human motion pattern, to form a biometric fingerprint that serves as a person's identifier [74] or other than to identity, to recognize gender via region-based clothing information in the case of insufficient face specification [73]; and appearance-based person re-identification in multi-camera surveillance systems [72]. There exists some work concerning semantic clothing attributes in fusion with vision features for surveillance purposes. Existing low-level features for person re-identification are complemented either by a full set of mid-level clothing attributes [9, 78] or by a selected and weighted subset of most-effective ones [3]. Whilst similar selection and weighting techniques enable optimised attributes [79] can be used in both automatic re-identification and identification [80].

In person re-identification problem, rather than relying only on single fame imagery features as most of current re-identification methods, it can be useful to learn a model for automatically selecting the most discriminative video fragments from noisy or incomplete image sequences as proposed by [81], where they investigate and discuss the effect of clothing variation on person re-id and consider it as a severe appearance change for vision-based features that might happen in practical surveillance scenarios. For their investigation, they evaluate their model and compare performance using different datasets embracing a proportion ranging from 0%-50% of people changing their clothing across cameras. They consider some cases as challenging, when clothing might be substantially changed in appearance within or across camera views, e.g. a worn jacket/coat that is later taken off and carried in hand instead. Such challenging exceptions or possible changes across non-overlapping cameras can be a likely future work to investigate the capabilities of our soft clothing biometric approaches and compare human-based with vision-based traits in similar surveillance scenarios.

1.5 Symbol Definitions

With a view to simplifying readability and understanding of the symbols used throughout the forthcoming identification and retrieval experiments, this section aims to list and briefly define all the denotation symbols used to describe a dataset (or gallery), as follows:

- **Clothing-based Soft Biometric Galleries**
  - ***Cat-21*** or ***ManCat-21*** — is a manual categorical-based gallery, which comprises 21 categorical soft clothing biometrics; describing all manually-annotated binary, multiclass, and relative clothing attributes.
  - ***Cat-7*** — is a manual categorical-based gallery, which comprises seven categorical soft clothing biometrics; describing only manually-annotated relative clothing attributes.
  - ***Cat-6*** — is a manual categorical-based gallery, which comprises six categorical clothing biometrics; describing selected manually-annotated clothing attributes, as the best correlated and most discriminative via ANOVA as a feature subset selection.

- ***Cmp*** or ***ManCmp-7*** — is a manual comparative-based gallery, which comprises seven comparative soft clothing biometrics; describing only manually-compared relative and comparable clothing attributes.
- ***AutoCat-17*** — is an automatic categorical-based gallery, which comprises 17 categorical clothing biometrics; describing all automatically-annotated clothing attributes.
- ***AutoCat-14*** — is an automatic categorical-based gallery, which comprises 14 categorical clothing biometrics; describing selected automatically-annotated clothing attributes, as the most discriminative via ANOVA as a feature subset selection.
- ***AutoCmp-10*** — is an automatic comparative-based gallery, which comprises ten comparative clothing biometrics; describing all automatically- compared relative clothing attributes.
- ***AutoCmp-5*** — is an automatic comparative-based gallery, which comprises five clothing biometrics; describing selected automatically-compared relative clothing attributes, as the most discriminative via ANOVA as a feature subset selection.
- ***Auto(Cat-14&Cmp-5)*** — is an automatic hybrid gallery, which is composed by fusing two automatic galleries *AutoCat-14* and *AutoCmp-5*; comprising 14 categorical and five comparative automatic clothing biometrics.

- **Body-based Soft Biometric Galleries**

  - ***tradSoft*** — is a manual categorical-based gallery, which comprises four categorical traditional soft biometrics; describing only manually-annotated global body attributes (i.e. Age, Ethnicity, Sex, and Skin Colour).
  - ***softBody*** — is a manual categorical-based gallery, which comprises (either 23 or 17) categorical soft body biometrics; describing all manually-annotated binary, multiclass, and relative body attributes, including the four global body attributes of *tradSoft*.

- **Combined Clothing and Traditional Soft Biometric Galleries**

  - ***tradCat-21*** — is a fusion of clothing (*Cat-21*) and traditional (*tradSoft*) galleries, which comprises 21 manual categorical soft clothing biometrics combined with 4 traditional soft biometrics.

- ***tradCat-7*** — is a fusion of clothing (*Cat-7*) and traditional (*tradSoft*) galleries, which comprises seven (relative) manual categorical soft clothing biometrics combined with 4 traditional soft biometrics.
- ***tradCat-6*** — is a fusion of clothing (*Cat-6*) and traditional (*tradSoft*) galleries, which comprises six manual categorical soft clothing biometrics (correlated and selected via ANOVA) combined with 4 traditional soft biometrics.
- ***tradCat-5*** — is a fusion of clothing (*Cat-5*) and traditional (*tradSoft*) galleries, which comprises five manual categorical soft clothing biometrics (selected via ANOVA) combined with 4 traditional soft biometrics.
- ***tradCmp*** — is a fusion of clothing (*Cmp*) and traditional (*tradSoft*) galleries, which comprises seven (relative) manual comparative soft clothing biometrics combined with 4 traditional soft biometrics.
- ***tradAutoCat-14*** — is a fusion of clothing (*AutoCat-14*) and traditional (*tradSoft*) galleries, which comprises 14 automatic categorical soft clothing biometrics (selected via ANOVA) combined with 4 traditional soft biometrics.
- ***tradAutoCmp-5*** — is a fusion of clothing (*AutoCmp-5*) and traditional (*tradSoft*) galleries, which comprises five automatic comparative soft clothing biometrics (selected via ANOVA) combined with 4 traditional soft biometrics.
- ***tradAuto(Cat-14&Cmp-5)*** — is a fusion of the automatic hybrid clothing gallery *Auto(Cat-14&Cmp-5)* and the traditional gallery *tradSoft*, which comprises 14 categorical and five comparative automatic clothing biometrics combined with 4 traditional soft biometrics.

- **<u>Combined Clothing and body Soft Biometric Galleries</u>**

- ***softCat-21*** — is a fusion of clothing (*Cat-21*) and body (*softBody*) galleries, which comprises 21 manual categorical soft clothing biometrics combined with (either 23 or 17) soft body biometrics.
- ***softCat-7*** — is a fusion of clothing (*Cat-7*) and body (*softBody*) galleries, which comprises seven (relative) manual categorical soft clothing biometrics combined with 23 soft body biometrics.

- ***softCat-6*** — is a fusion of clothing (*Cat-6*) and body (*softBody*) galleries, which comprises six manual categorical soft clothing biometrics (correlated and selected via ANOVA) combined with 17 soft body biometrics.
- ***softCat-5*** — is a fusion of clothing (*Cat-5*) and body (*softBody*) galleries, which comprises five manual categorical soft clothing biometrics (selected via ANOVA) combined with 23 soft body biometrics.
- ***softCmp*** — is a fusion of clothing (*Cmp*) and body (*softBody*) galleries, which comprises seven (relative) manual comparative soft clothing biometrics combined with (either 23 or 17) soft body biometrics.
- ***softAutoCat-14*** — is a fusion of clothing (*AutoCat-14*) and body (*softBody*) galleries, which comprises 14 automatic categorical soft clothing biometrics (selected via ANOVA) combined with 17 soft body biometrics.
- ***softAutoCmp-5*** — is a fusion of clothing (*AutoCmp-5*) and body (*softBody*) galleries, which comprises five automatic comparative soft clothing biometrics (selected via ANOVA) combined with 17 soft body biometrics.
- ***softAuto(Cat-14&Cmp-5)*** — is a fusion of the automatic hybrid clothing gallery *Auto(Cat-14&Cmp-5)* and the body gallery *tradSoft*, which comprises 14 categorical and five comparative automatic clothing biometrics combined with 17 soft body biometrics.

1.6 Outline of the Thesis

The thesis structure comprises seven more chapters, as follows:

- **Chapter 2** describes the new semantic clothing attributes and descriptive labels for annotation, human based clothing label data acquisition, and resulting soft clothing biometrics in categorical and comparative forms and the methodology used to derive these soft biometrics [P1].
- **Chapter 3** briefly builds a common understanding of the significance, differences, correlations, and efficacy of soft clothing traits in identification, through a series of statistical analysis methods [P2, P5].

- **Chapter 4** experimentally investigates the power and evaluates the performance of soft clothing biometrics in human identification in case of addition to other soft biometrics or when used alone [P1].

- **Chapter 5** presents the use of soft clothing biometrics for enhanced subject retrieval and how they can be generalized for unseen data [P2]. It also studies the utility of soft clothing traits in more challenging biometric retrieval in realistic scenarios and examines their capability in achieving viewpoint invariant subject retrieval [P3].

- **Chapter 6** provides a brief description of proposed automatic soft clothing attributes and explores the methodology used for vision-based automatic extraction and annotation. It then presents the ensuing automatic soft clothing biometrics and the detailed analysis of these traits. Eventually it demonstrates a number of experiments on viewpoint invariant retrieval, investigating the power of automatic soft clothing traits derived via computer-vision and allowing performance comparison against those manual soft clothing traits inferred by humans [P4].

- **Chapter 7** studies the clothing feature space via extended analysis and empirical investigation of the capabilities of soft biometrics using clothing attributes in human identification and retrieval, leading to an insightful guide for feature subset selection and enhanced performance [P5].

- **Chapter 8** concludes the thesis by summarising the main results and findings and outlining future research lines.

Chapter 2

Soft Clothing Biometrics

2.1 Semantic Clothing Attributes

There are many possible clothing attributes and labels; we study an initial set of those which appear to be basic. Based on a number of general considerations, a set of attributes and their labels is chosen and defined so as to construct a feature vector. Firstly, this set of defined attributes is intended to be structural and comprehensive rather than a detailed description of clothing. The set is also suited to analysis of surveillance data, which is often vulnerable to challenges and obstacles such as low resolution, long distance, lighting conditions, pose, occlusion, and change of viewpoint [1, 8, 11, 27, 33, 39]. Besides that every single trait may have its own challenges, which are related to it but not to the others [30]. Secondly, the attributes are desired to be as understandable as possible, reliable, visible, and observable, allowing an annotator to describe what is being seen not what is being inferred. By taking these considerations into account a number of attributes deemed suitable for later investigation were excluded. Here we propose a set of semantic clothing attributes described with *categorical* and *comparative* labels.

Each categorical label, which is designated for a non-relative attribute, is assigned an integer value as a numeric representation for its textual expression. Previously a practical psychological analysis [82] using bipolar scales to define traits was performed as a unique analysis on whole-body descriptions. We utilize this validated practical mechanism to better reflect the degree-of-strength of the relative attributes. Each of the seven relative attributes is formulated as a bipolar five-point scale. We assign a set of ordered categorical labels ranging from 1 to 5 and a set of comparative labels is assigned accordingly ranging from -2 to 2. For all binary-label clothing attributes, such as 'Belt presence', we add a label 'Don't know' as a choice, which often can be a possible reasonable answer in real-life, reflecting the user's uncertainty as to whether an attribute is present or not.

Table 2.1: Semantic clothing attributes and corresponding categorical and comparative labels

Body zone	Semantic Attribute	Categorical Labels	Comparative Labels
Head	1. Head clothing category	[None, Hat, Scarf, Mask, Cap]	
	2. Head coverage	[None, Slight, Fair, Most, All]	[Much less, Less, Same, More, Much more]
	3. Face covered	[Yes, No, Don't know]	[Much less, Less, Same, More, Much more]
	4. Hat	[Yes, No, Don't know]	
Upper body	5. Upper body clothing category	[Jacket, Jumper, T-shirt, Shirt, Blouse, Sweater, Coat, Other]	
	6. Neckline shape	[Strapless, V-shape, Round, Shirt collar, Don't know]	
	7. Neckline size	[Very small, Small, Medium, Large, Very large]	[Much smaller, Smaller, Same, Larger, Much larger]
	8. Sleeve length	[Very short, Short, Medium, Long, Very long]	[Much shorter, Shorter, Same, Longer, Much longer]
Lower body	9. Lower body clothing category	[Trouser, Skirt, Dress]	
	10. Shape (of lower clothing)	[Straight, Skinny, Wide, Tight, Loose]	
	11. Leg length (of lower clothing)	[Very short, Short, Medium, Long, Very long]	[Much shorter, Shorter, Same, Longer, Much longer]
	12. Belt presence	[Yes, No, Don't know]	
Foot	13. Shoes category	[Heels, Flip flops, Boot, Trainer, Shoe]	
	14. Heel level	[Flat/low, Medium, High, Very high]	[Much lower, Lower, Same, Higher, Much higher]
Attached to body	15. Attached object category	[None, Bag, Gun, Object in hand, gloves]	
	16. Bag (size)	[None, Side-bag, Cross-bag, Handbag, Backpack, Satchel]	[Much smaller, Smaller, Same, Larger, Much larger]
	17. Gun	[Yes, No, Don't know]	
	18. Object in hand	[Yes, No, Don't know]	
	19. Gloves	[Yes, No, Don't know]	
General style	20. Style category	[Well-dressed, Business, Sporty, Fashionable, Casual, Nerd, Bibes, Hippy, Religious, Gangsta, Tramp, Other]	
Permanent	21. Tattoos	[Yes, No, Don't know]	

2.1.1 Categorical Clothing Labels

Categorical labels can be defined as nameable descriptions used to describe semantic attributes of an individual's clothing, usually associated with multiple clothing categories or styles such as (*Upper body clothing category*: 'Jacket', 'Jumper', 'T-shirt' etc.) or can be labels describing the degree of presence of relative attributes such as (*Sleeve length*: 'Very short', 'Short', 'Medium', 'Long', 'Very long').

2.1.2 Comparative Clothing Labels

Comparative labels are nameable descriptions used to describe only relative attributes of an individual's clothing compared with another individual's clothing. In other words, these labels describe the degree of comparisons of relative attributes, such as (*Neckline size*: 'Much smaller', 'Smaller', 'Same', 'Larger' and 'Much larger').

We propose a list of 21 semantic attributes and for each attribute, a suitable group of categorical labels is specified to be used for describing these attributes. Furthermore, seven of the aforementioned 21 attributes are both categorical and relative, whereas the remaining 14 are unsuited for comparison because they are binary or multi-class attributes that can be described using only categorical (absolute) labels. Thus, the categorical labels of these seven relative attributes are extended to their corresponding comparative labels. Table 2.1 shows the list of proposed semantic attributes with their assigned categorical and comparative labels, where the seven relative comparable attributes are in bold.

2.2 Human Clothing Dataset

The Soton Gait Database (SGDB) [83] is a standard database used for this research, as it comprises a subset of full-body front- and side-view still images of each subject in the database with soft biometrics' labels available [27, 39]. The front-view images are used to collect the clothing descriptions. This subset consists of 115 individuals with a total of 128 front samples. Each sample of each individual is handled, throughout all the steps of clothing-based experiments, as an independent individual. That is due to the fact that all individuals with multiple samples are wearing different clothing in each sample. Hence, multiple samples of a single individual are considered as different and independent entities by which each entity represents that individual if wearing exactly the same clothing. Otherwise, it is considered as another entity, even though it belongs to the same individual, but wearing different clothing. In this research we refer to each entity (i.e. sample in this dataset) as a ***subject***. Note that about 90% of the subjects in the database wear largely similar clothing (jeans, T-shirt, etc.), which appears sufficiently challenging for this initial study.

2.2.1 Body Description Labels

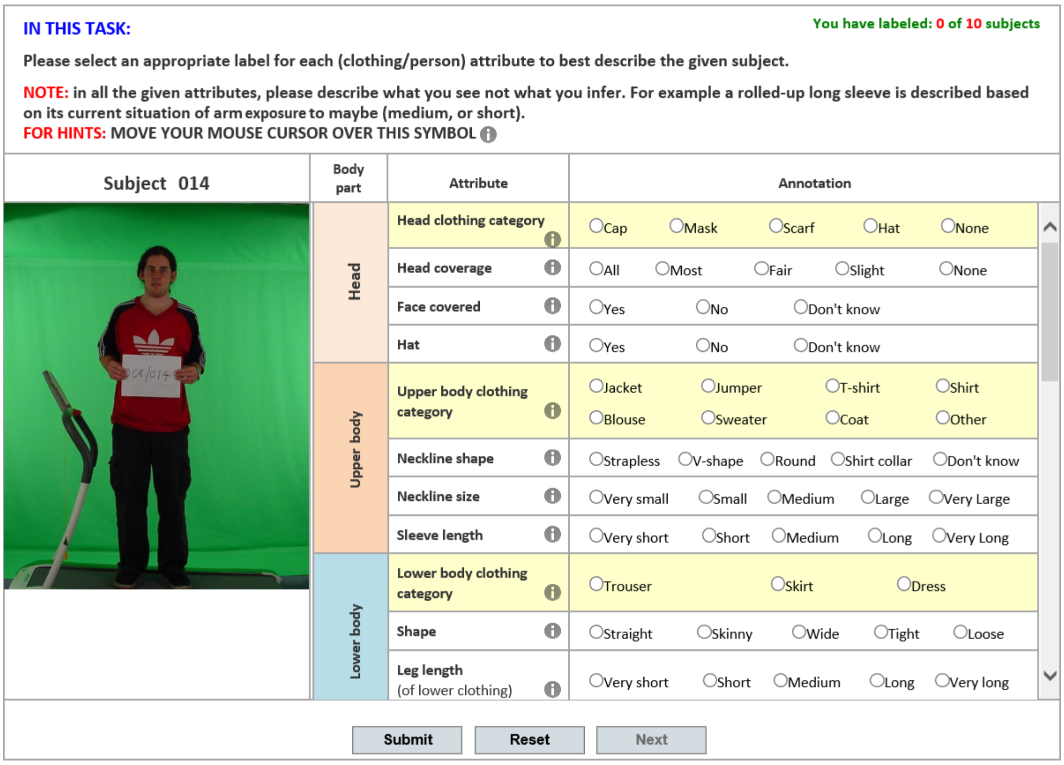
Soft body descriptions are available for Soton database [83], which were previously proposed and used in categorical form [39] then extended later to comparative form [27]. Categorical soft body descriptions [39] were collected such that, each of 115 individuals was labelled by multiple users' describing 23 soft body traits. Table 2.2 shows these body traits and associated semantic terms, which they were grouped into three categories: *Body shape*, *Global*, and *Head*. This body description data is used in this research in a variety of ways and purposes, such as a baseline for performance comparison, and as traditional soft traits to be supplemented by new soft clothing traits.

Table 2.2: Soft body traits and associated semantic terms [39]

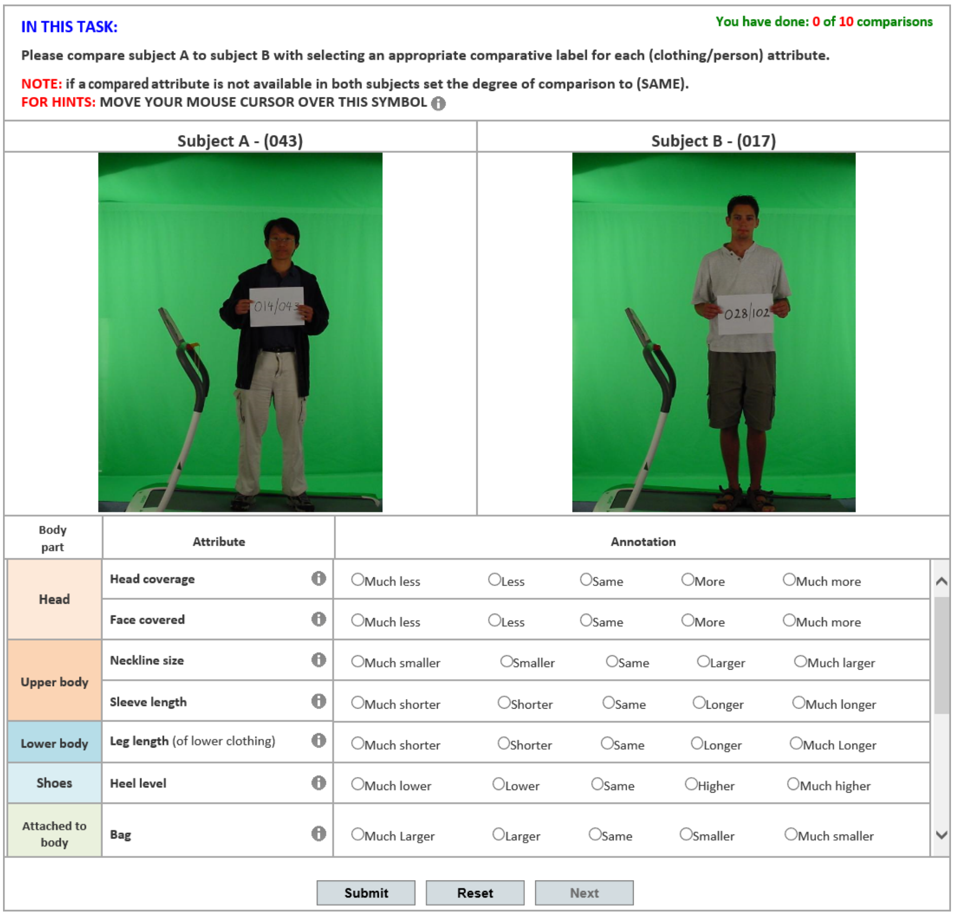
Group	Body trait	Terms
Body shape	1. Arm length	[Very Short, Short, Average, Long, Very Long]
	2. Arm thickness	[Very Thin, Thin, Average, Thick, Very Thick]
	3. Chest	[Very Slim, Slim, Average, Large, Very Large]
	4. Figure	[Very Small, Small, Average, Large, Very Large]
	5. Height	[Very Short, Short,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

2.3 Clothing Description Data Acquisition

A web-based Clothing Annotation Procedure (CAP) was designed and developed to obtain clothing labels and comparisons, as shown in Fig. 2.1. The annotation procedure was split into two tasks. The first task required a user to annotate ten subjects. Each subject was described by selecting 21 appropriate categorical labels 2.1-a. The second task required a user to compare one subject, selected randomly from the ten already annotated, with other ten new subjects. A comparison, between two subjects was performed by selecting seven appropriate comparative labels 2.1-b. In this way M labellers provided N labels on O subjects. With a view to simplifying and organizing user annotations, the clothing attributes were grouped based on their zones and relevance as follows: Head, Upper body, Lower body, Foot, Attached to body, General style and Permanent as shown in Table 2.1, which was similarly used to view attributes and labels in annotation forms as illustrated in Fig. 2.1.



(a) Categorical annotation form of Task 1



(b) Comparative annotation form of Task 2

Figure 2.1: The CAP website developed to obtain labels and comparisons

The annotation web-forms were carefully designed to take into account a number of potential psychological factors, which were also considered earlier in collecting soft body labels [39], as they include:

- *Anchoring*, which could occur when asking a question with some initial default answer a person's answer may often be anchored around those initial answers [84]. Therefore, to avoid this we left all labelling options (radio buttons) initially unchecked, forcing a pure response from the annotator.
- *Memory*, as the time passage could affect the recall of a subject's traits by a (witness) person [85]. Therefore, we allowed the annotator to see the described subject during the whole time of annotation to avoid the dependence on memory, assuring a confident description from the annotator to what is seen not what is remembered.
- *Observer variables*, which could affect a person's judgment of physical variables (measurements), since they often analyse other people in comparison with themselves, namely based on their perception to their own measurements [86, 87]. However this factor has less impact on clothing traits than body traits, since the annotator tends to focus more on clothes types and appearance. Therefore we emphasized repeatedly to the annotator that the objective is to describe a subject's clothing measurements relative to the subject's body (not the body itself), avoiding the annotator bias and confusion caused by the body measurements (e.g. confusing the trouser leg length with the subject leg length).
- *Defaulting*, which could occur when a person skips some descriptions since they assume them as known information by default [88]. Therefore, to avoid this problem we required responses to all questions as so the annotator could not leave out any description, forcing a response from the annotator for every single attribute even if it is label could be known (by default) from those other attributes already given by the annotator. For example, if the upper clothing was labelled by the annotator as "Jacket" they may leave out the sleeve length, since it might be considered by their defaulting perception as "long", while we still give some room of more discrimination (by specifying whether it is "Long" or "Very long"), and also allow for some possibility of exceptions (e.g. rolled up sleeve) or unusual fashions by giving the chance for the annotator to carefully choose the best descriptive label (which could be "Short" or "Medium") based on what is exactly seen not what is

inferred. Furthermore, we addressed some obvious potentials of annotators' defaulting by enforcing our controlled defaulting (i.e. applying auto label selection) in some definite cases when the selection of a particular label consequently and compulsory leads to another label (e.g. if head coverage labelled as "None" the hat presence definitely has to be "No"), which not only avoided the annotator's defaulting but also confliction between labels.

Categorical and comparative labels were collected from 27 users via the CAP website. All 128 samples were labelled by multiple users, with one or more separate user annotations per subject describing the 21 categorical attributes. All subjects were compared using the seven relative attributes by multiple users. To enrich the comparison data from the available number of collected comparisons, additional comparisons were inferred when two subjects were both compared with another same subject. A summary of collected user data and the inferred comparisons data is shown in Table 2.3.

Table 2.3: The number of collected and inferred user data

Data summary	Collected	Inferred	Total
Total user annotations	316	N/A*	316
Total user comparisons	317	556	873
Total categorical labels	6636	N/A	6636
Total comparative labels	2219	3892	6111
Average user annotations per subject	2.47	N/A	2.47
Average user comparisons per subject	2.48	4.34	6.82
Average categorical labels per subject	51.84	N/A	51.84
Average comparative labels per subject	17.34	30.41	47.75

* N/A= not applicable

2.4 Ranking by Relative Clothing Attributes

The attribute-based clothing comparison data cannot be used for identification, unless it is arranged as a list of subjects ordered with respect to a single attribute. Subsequently, the attributes of each subject can be described by relative measurements derived from per attribute ordering data. Different ranking methods can be applied for this purpose, such as the Elo ranking system [27] or a Ranking SVM [31, 89]. In this research, to achieve ordering and to derive the desired relative measurements, a soft-margin Ranking SVM method [90] is used, along with a supporting formulation of similarity constraints [31]. We decided to use Ranking SVM rather than using Elo ranking for a number of reasons including: Ranking SVM is well established and more standard in ranking problems [90]; It is more robust method compared with Elo in terms of the capability to achieve a desired ordering even with very small number of all possible comparisons [31, 89]; and learning a Ranking SVM function is applied only once during the training phase then learnt function can be used further to apply ranking for a new sample with no need for re-training [89], which is much less expensive than Elo rating that is excessively iterative-method and need to apply rating every time to rank a new sample [27].

We use Ranking SVM method to apply a pairwise technique based on learning a ranking function per attribute (e.g. *Sleeve length*). This ranking function is able to explicitly enforce a desired ordering on the training subject samples (i.e. in case of *Sleeve length* attribute, all subjects are ordered starting from those wearing the longest sleeves and ending with those wearing the shortest sleeves). Fig. 2.2 clarifies the concept of learning a ranking SVM function (w_m) and differentiate it from learning a typical SVM classifier (w_b). In the ranking example (on the right), the margin is defined as the distance between the closest two projected points among all projections (1-6), which enforces ordering on all training points. In the binary classifier example (on the left), the margin is defined as the distance between the closest two projected points (one from each class), which separates the two classes (+ and –) in such a way does not necessarily maintain ordering of all training points.

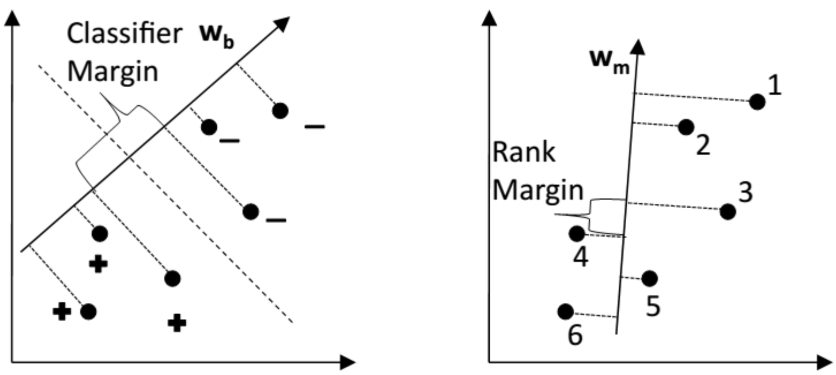


Figure 2.2: The difference between learning a wide-margin a ranking SVM function (right) and learning a wide-margin binary SVM classifier (left) [31]

Such learned ranking functions can be used not only to perceive the relative strength of attributes in a training sample, but also to predict the relative strength in a new test sample. Thus, for a set of attributes A, a ranking linear function r_a is learned for each attribute a such that:

$$r_a\,(x_i\,) = w_a^T\,x_i \tag{2.1}$$

where w_a is the coefficient of the ranking function r_a and x_i is a feature vector of attributes of a 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 $(i, j) \in D_a \Rightarrow i \succ j$, namely subject i has stronger presence of 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 similar presence of 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:

$$\begin{aligned} &\text{minimise} \quad \left(\frac{1}{2} \parallel w_a^T \parallel^2 + C \sum \xi_{ij}^2\right) \\ &\text{subject to } w_a^T (x_i - x_j) \geq 1 - \xi_{ij}\,; \quad \forall (i, j) \in D_a \\ &\qquad | w_a^T (x_i - x_j) | \leq \xi_{ij}\,; \quad \forall (i, j) \in S_a \\ &\qquad \xi_{ij} \geq 0 \end{aligned} \tag{2.2}$$

The degree of misclassification is measured by ξ_{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 using Eqn. (2.1) to a corresponding feature vector comprising a number of real-value relative measurements. Each measurement represents the relative strength of a single attribute.

2.5 Soft Clothing Traits

2.5.1 Categorical Clothing Traits (*Cat-N*)

Categorical labels are used to form two feature vectors for each subject in the dataset. In both versions of feature vectors, an average label per attribute is calculated for a set of labels provided by multiple users describing the same subject. The first feature vector is formed from 21 categorical normalized average-labels, describing all the 21 clothing attributes to build the first gallery of categorical clothing traits for all subjects (referred to as *Cat-21*). Note that the symbol *Cat-21* and all forthcoming symbols of biometric galleries are listed and defined earlier in (page 14-16), under Section 1.5. Similarly, the second feature vector is formed from a subset of only the seven relative clothing attributes to build the second gallery (referred to as *Cat-7*).

A multi-attribute soft biometric signature can be derived from a set of categorical labels per attribute for each subject via different methods such as the mean (or average) label [27, 29, 39], the median label [27, 41, 50], the mode (or the most frequent label) [91], and even by simply considering every multi-attribute annotation (provided by a single annotator) as an (individual) biometric signature of a subject; as such since a subject may have multiple annotations, this implies a subject may have multiple biometric signatures. Amongst those possible methods, we found that an average-label (or mean-label) per attribute as the most effective representation to compose a single discriminative soft biometric signature for a subject in database. We then normalized each average-label per attribute α by subtracting the overall mean (μ_α) and divided by the overall standard deviation (σ_α) of that attribute (α) across all samples in the gallery, such that a new normalized average-label $\ell'_{s,\alpha}$ for subject s, describing and attribute α is deduced as $\ell'_{s,\alpha} = (\ell_{s,\alpha} - \mu_\alpha)/\sigma_\alpha$.

2.5.2 Comparative Clothing Traits (*Cmp*)

Comparison data can be used to convey meaningful information describing a subject in relation to the remaining population [27]. The collected comparative annotations need to be anchored, per attribute, to define invariant relative measurements for each subject. To derive these measurements, a ranking method needs to be applied to arrange a list of ordered subjects with respect to a single attribute. In the ranking process, the comparisons between subjects are used as rules to enforce ordering for subjects, and to adjust a relative measurement per attribute for each.

To perform identification using comparative labels, the corresponding relative measurements are derived using the Ranking SVM method (described in Section 2.4). All subjects in the *Cat-7* gallery are used as a training dataset to learn seven optimal ranking functions for the seven relative attributes. The coefficient w is derived using the formulation in Eqn. (2.2). The desirable per attribute ordering of all subjects is derived using w. Then by Eqn. (2.1), each w is used to map each feature vector in *Cat-7* to a corresponding vector of seven relative measurements (i.e. comparative traits) describing a single subject. Hence, all the obtained relative measurement vectors are gathered to compose a new gallery of comparative traits (referred to as *Cmp*). Table 2.4 summaries the proposed categorical and comparative soft clothing traits (galleries).

Table 2.4: Proposed soft clothing traits and galleries

Trait	Description
Cat-21	21 categorical normalized average-labels, describing 14 binary/multiclass and 7 relative attributes
Cat-7	7 categorical normalized average-labels, a subset of ***Cat-21*** describing only the 7 relative attributes
Cmp	7 comparative traits, mapping ***Cat-7*** to corresponding relative measurements based on comparisons per each of the 7 relative attributes

2.6 Methodology Framework

We describe our exploratory experimental approach designed to obtain and analyse label and comparison data, which is used to build a soft biometric database of clothing descriptions and to investigate the significance and correlations of the proposed attributes, towards achieving and evaluating human identification and retrieval. Fig. 2.2 represents a high-level summary of our methodology used along this research, illustrated as a framework, which is partly detailed in the preceding of this chapter and will be progressively accomplished in forthcoming experimental chapters 4 and 5.

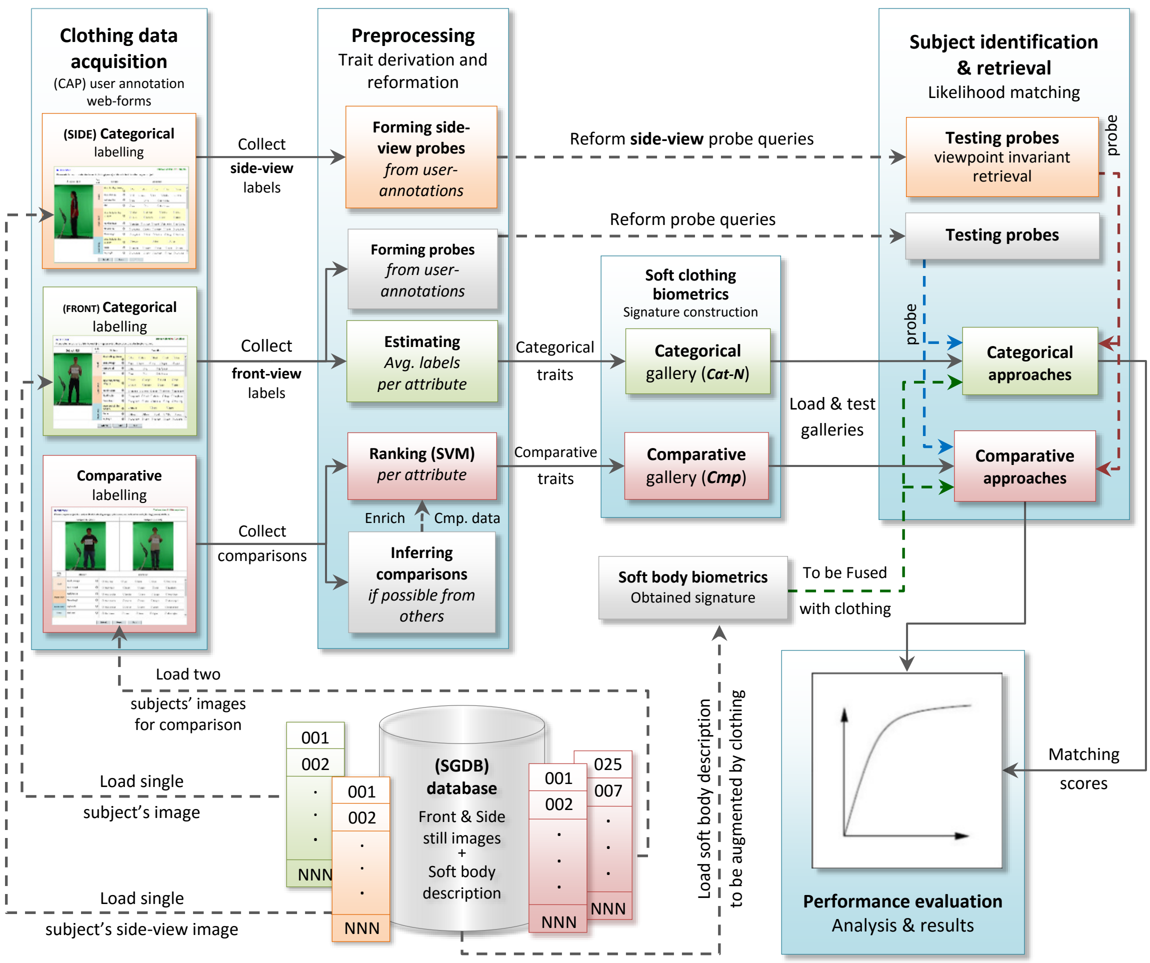


Figure 2.3: Methodology framework of the current research approach

This framework comprises four main stages. The first stage denotes clothing description data collection via CAP website, where still subject images are loaded from SGDB database and randomly displayed for human-based annotation. Two web-forms are associated with categorical labelling; one using a single front view image to be annotated for primary identification and the other using a single side view image to be annotated for more challenging retrieval tasks. A third web-form concerns pairwise comparative labelling for a subject against another.

The second stage shows preprocessing methods applied to collected labels and comparisons for trait derivation and feature vectors reformation for both gallery and probe sets. Thus a normalized average labels per attribute is estimated for each subject creating a biometric signature for them. Testing probes are randomly selected from available front and side categorical user annotations and reformed according to each examined gallery to be suited for matching and evaluation. For comparative clothing traits, attribute comparisons between subjects are firstly enriched by inferring all possible comparisons from available ones then the Ranking SVM method is used to derive relative measurements to be used as comparative clothing traits. Consequently suitable testing probes are derived and mapped to corresponding relative measurements.

The third stage represents construction of all categorical and comparative galleries of soft biometric signatures used for identification and retrieval purposes. Finally the fourth stage outlines clothing-based subject identification and retrieval experiments using different proposed soft biometric approaches, leading to performance evaluation and analysis of soft clothing traits in supplementing soft body biometrics and further when used alone.

2.7 Conclusions

In this chapter the new set of proposed semantic clothing attributes is introduced along with their descriptive categorical and comparative labels. The database adopted for this research is described and the label and comparison data acquisition via the clothing annotation website (CAP) is explained. Soft body descriptions, available in the database (SGDB), are briefly reviewed since they will be used as a baseline for performance comparison, and as traditional soft traits to be supplemented by the new soft clothing traits. The resulting soft clothing biometrics are defined in their both categorical and comparative forms, where the Ranking SVM used for deriving usable comparative traits is explained. Eventually an overview is provided as a methodology framework, summarizing all stages and components of this chapter and further additional relevant stages and steps in the following chapters.

Chapter 3

Soft Clothing Data Analysis

3.1 Attribute Significance and Correlation

The exploration of attributes' significance and correlations, is deemed to be an important analysis resulting in a better comprehension of which of attributes contribute most to identification and leading to wider potential predictability of other attributes [43]. We assessed the proposed clothing attributes to investigate their correlation and effectiveness for subject description. Whilst different possible methods could be used to investigate attribute significance, we utilized well established statistical analysis for this investigation. We used (Pearson's *r*) coefficient to compute the correlation and derive the correlation matrix highlighting the significance of attributes and the mutual relations between attributes, which can be computed as follows:

$$r = \frac{\sigma_{XY}}{\sigma_X \, \sigma_Y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{3.1}$$

where *X* and *Y* are two variables representing the label values of two different semantic traits simultaneously used to describe an individual. Namely, X_i and Y_i are two labels of i^{th} annotation describing the same subject. So, *r* is calculated by dividing the covariance of *X* and *Y*, denoted as σ_{XY}, by the product of σ_X and σ_Y the standard deviations of *X* and *Y* respectively. The resulting value of *r* ranges from -1 to 1 to represents the correlation coefficient between two labels. When the value of *r* is 1, it means that *X* and *Y* are positively correlated and their all data points lie on a straight line. On the other hand, a value of *r* = -1, this implies that *X* and *Y* are also correlated but negatively. A value of *r* = 0 indicates that there is no correlation between *X* and *Y*. Hence, the closer *r* is to 1 reflects a higher positive relationship between labels. On the other hand, the closer *r* is to -1 reflects a higher *negative* relationship between labels. Note that, when one or both of correlated labels are binary or multi-class describing pure (nominal) categorical attributes, this indicates that they are simultaneously present in a single annotation. That is because the assigned numeric values of such labels do not reflect ordering or measurements, unlike (ordinal) labels of relative attributes, so they can be assigned values in any order or their assigned values can be consistently exchanged.

We generated several correlation matrices of all possible correlations for the available collected labels, comparisons and further derived traits. The calculated linear correlation coefficient was considered as significant with respect to the resulting p-value when $p \leq 0.05$. It is worth noting that the low correlation between two different attributes does not suggest that there is no relationship between the two attributes, but conveys a notion that this correlation is not prevalent within the clothing dataset currently used such as the attributes 'Face covered' and 'Bag'.

3.1.1 Correlations between Soft Clothing Labels

We study the proposed clothing attributes' relationships using the (Pearson's r) correlation matrix. We generated correlation matrices for the available collected labels, comparisons and further derived traits. The calculated correlation coefficient was considered as significant when the resulting p-value when $p \leq 0.05$. Fig. 3.1 demonstrates the correlation between all labels of 18 clothing attributes (see Table 2.1); attributes without correlation are not shown, since their labels have very low variability as they are not prevalent in the dataset such as '*Face covered*' (3) and '*Bag*' (16).

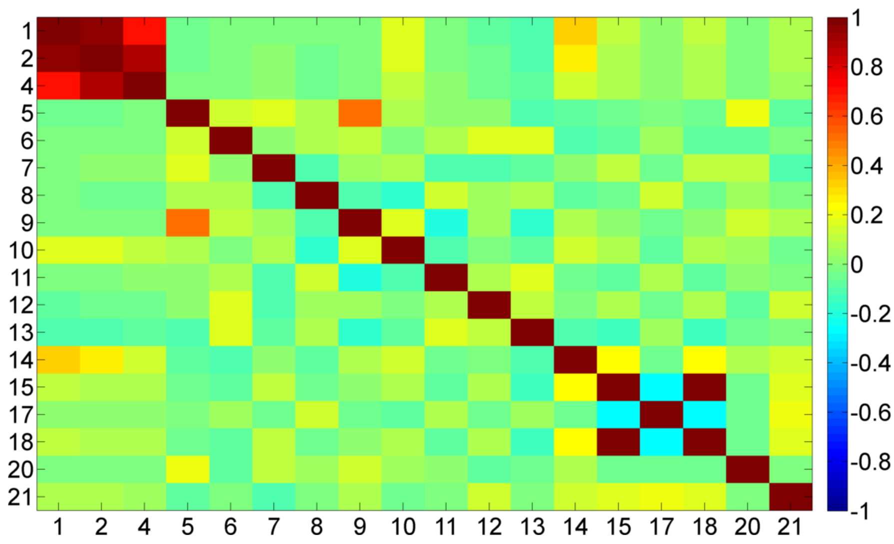
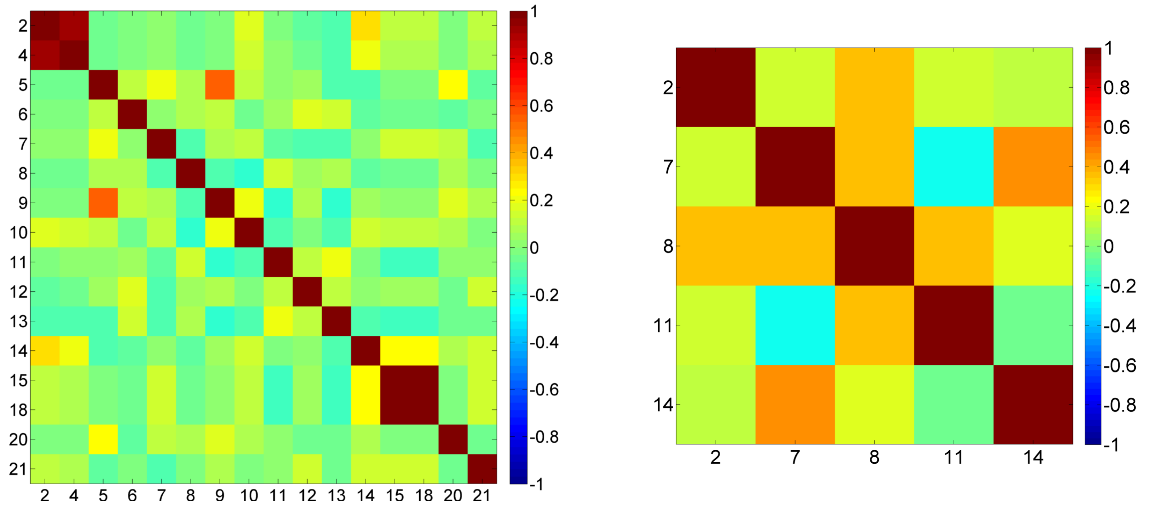


Figure 3.1: Clothing label correlation matrix

High correlation is symbolized by orange, and low by blue/ green. As such, attributes relating to head coverage are highly correlated, as are the attributes (15) and (18) relating to the description of items attached to the body. Clothing is well correlated for upper (5) and lower (9) body. The matrix structure suggests that the desired uniqueness has been achieved.

3.1.2 Correlations between Derived Soft Clothing Traits

The proposed categorical and comparative clothing traits were assessed to investigate their correlation and effectiveness for subject description. We generated correlation matrices for both categorical and comparative traits, which were derived from label data as normalized average labels and from comparison data as relative measurements, as described in Section 2.5. The calculated correlation coefficient was considered here also as significant when the resulting p-value when $p \leq 0.05$.



(a) Categorical clothing traits

(b) Comparative clothing traits

Figure 3.2: Correlations between soft clothing traits

Fig. 3.2 demonstrates the correlation between the most significant categorical and comparative traits (see Table 2.1); traits without correlation are not shown, due to the absence or rarity these traits (or biometrics) in the dataset, which are derived from collected labels such as '*Gloves'* (19) and '*Head coverage*' (2). High correlation is symbolized by orange, and low by blue/green. In the categorical matrix, traits relating to head coverage (2) and (4) are highly correlated, as are the traits (15) and (18) relating to the description of items attached to the body. Clothing categories are well correlated for upper (5) and lower (9) body, as expected. In the comparative matrix, neckline size (7) and heel level (14) are well correlated. Neckline size (7) is negatively correlated with leg length (11). The structures of both correlation matrices suggest that the desired uniqueness has indeed been achieved.

3.2 Clothing Trait Distribution Statistics

Studying and analysing clothing trait distributions is useful not only to understand the nature of the population of these traits but also to enable comparison between clothing traits from a statistical point of view. Fig. 3.3 and 3.4 demonstrate the population of each categorical and comparative soft clothing traits and clarify the differences in their data distributions. It can be perceived that some categorical traits such as *'Leg Length (of lower clothing)'* (11) and '*Neckline Size*' (7) appear to be more fitted to the normal distribution structure, whereas some others are less fitted such as 'Shoes category' (13) and '*Shape (of lower clothing)*' (10).

Thus, it is not necessarily for each clothing trait (in soft biometric signature) to be well fitted to normal distribution in order to be a valid soft biometric and some traits are expected to be more compliant to normal distribution than some others. This is in contrast to soft body traits, which are expected to offer more normal data distribution than clothing [92], since body attributes are natural and inherent, whereas clothing attributes are artificial and behavioural. This is due to the fact that populations of clothing attributes, (the presence/absence, and the prevalence/rarity) are more liable to many psychological, cultural, social, environmental, seasonal, and geographical factors, whilst this observation was quite similarly indicated in [10, 64]. For instance, a particular upper/lower clothing category can be predominant because of a certain seasonal factor like wearing coats in winter or shorts in summer. However, relative soft clothing traits appear to be less affected by such factors. For example, a relative attribute, describing sleeve/leg length will remain most likely operable, distinguishable, and comparable, even though for people wearing very similar upper/lower clothes under the effect of any of those factors like people wearing suits in business or shirts in university.

Based on the obtained results of this analysis of soft clothing information, it is worth noting the following findings: regardless the type of soft clothing traits (categorical or comparative), the absence or rarity of a particular trait in a database will consequently affect the data distribution of that trait e.g. '*Head coverage*' (2) and '*Gloves'* (19); comparative traits often tend to display more normal distribution than their categorical counterparts, since they are represented by relative measurements derived using the comparative labels as explained previously in Section 2.4; and the categorical label data distributions of the binary/multi-class traits lean to be either unlikely or the least to satisfy the normal distribution assumption, as expected.

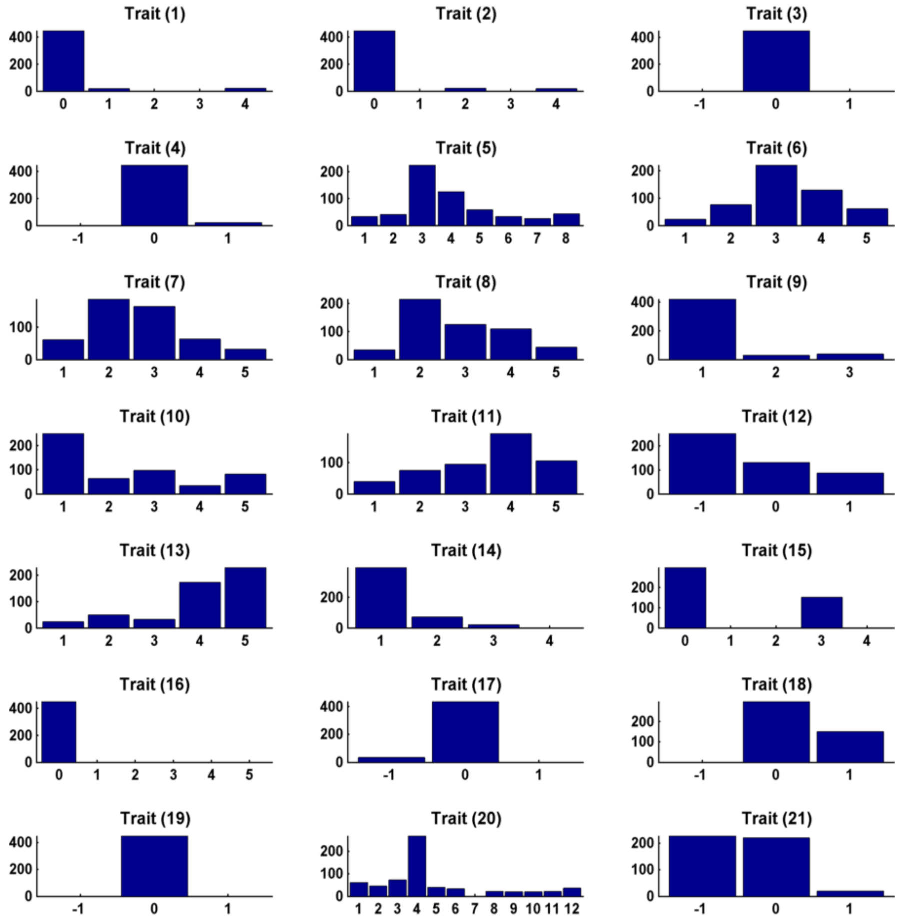


Figure 3.3: Data distribution of categorical soft clothing traits

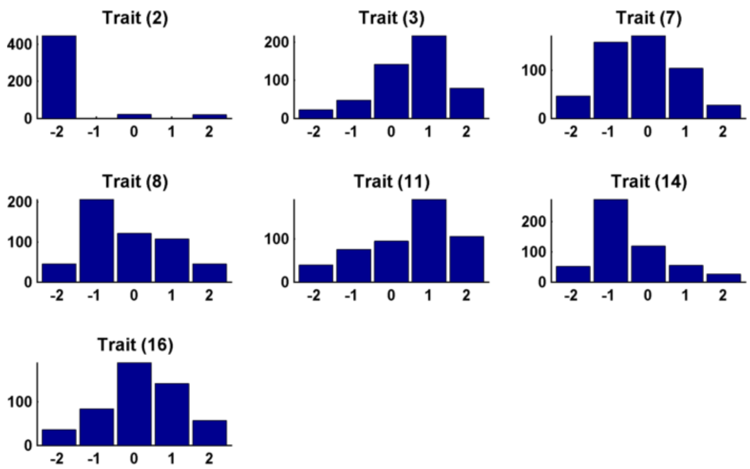


Figure 3.4: Data distribution of comparative soft clothing traits

3.3 Analysis of Variance (ANOVA) for Clothing Traits

ANOVA is a widely used statistical test, and has been effectively used in several investigational studies. In our study, a one-way ANOVA is used to analyse the variance for every single soft clothing trait, which is produced as a result of multiple annotation labels collected from multiple annotators describing the same subject.

The concept of this analysis process, known as the F-test, relies on analysing a single variable described by multiple observations, and these observations underlying two or more independent groups. This, in turn, allows comparison between intra-variance *within* each of all groups with the mean inter-variance *between* these groups, resulting in a value defined as F-ratio. A further value can be deduced form F-ratio and the F-distribution known as p-value, which is supposed to be less than or equal to a certain small value (usually 0.05 or 0.01), to consider F-value as significant and to reject the null-hypothesis [93].

The computed F-ratio and its corresponding p-value are used to order the categorical and comparative soft clothing biometrics by their estimated capability of distinguishing between subjects. A larger F-ratio with a smaller p-value is better, where the total degree of freedom (*df*) = 443 for F-ratio, and $p \leq 0.05$ for significance level. F-ratio is calculated as follows:

$$\text{F-ratio} = \frac{\sum_i n_i(\bar{X}_i - \bar{X})^2/(K-1)}{\sum_{ij}\left(X_{ij} - \bar{X}_i\right)^2/(N-K)} \tag{3.2}$$

The upper bound in this Eqn. (3.2) yields the (total *between-groups* variance), whereas the lower bound yields the (total *within-group* variance). In our biometric context, each group represents a subject, where K is the number of subjects and N the total number of samples of all subjects, while n_i represents the number of sample of i^{th} subject. X_{ij} denotes the value of j^{th} sample of the i^{th} subject, $\bar{X}_i$ denotes the mean across all samples of the i^{th} subject, and $\bar{X}$ is the mean across all N samples. The degree of freedom is enforced by $K-1$ on subject level and $N-K$ on sample level.

The reported results show the most significant soft clothing traits, and likely suggest which traits are more successful than others. Table 3.1 provides the ordered lists of resulting ANOVA test values for categorical and comparative clothing traits. Lower body clothing category (9) is observed as the most discriminative trait. Head coverage (2) is highly discriminative since few subjects had covered heads. It is perhaps surprising that sleeve length (8) is so discriminative, especially compared with the length of the trousers (11), but that is what this analysis reveals, and no summary analysis is possible by human vision. It is worth applying further multivariate analysis of variance (MANOVA) on the whole set of clothing attributes, which explores interactions between multiple variables (attributes) and measures significance of multivariate discrimination capability via different standard statistics [93], as shown in Table 3.2, where all measures provide significant p-value <0.0001.

Table 3.1: ANOVA results of soft clothing traits ordered by p-value

(a) Categorical traits

Clothing attribute	**F-ratio** (df = 443)	**p-value** ($p \leq 0.05$)
9. Lower body clothing category	19.747	2.02E-99
2. Head coverage	19.651	3.90E-99
8. Sleeve length	12.590	3.60E-74
1. Head clothing category	11.054	2.46E-67
11. Leg length (of lower clothing)	5.040	5.48E-32
4. Hat	4.913	4.87E-31
5. Upper body clothing category	4.738	1.03E-29
12. Belt presence	4.457	1.52E-27
13. Shoes category	4.282	3.64E-26
6. Neckline shape	2.413	1.68E-10
20. Style category	2.349	5.83E-10
14. Heel level	1.863	5.58E-06
7. Neckline size	1.688	1.15E-04
10. Shape (of lower clothing)	1.626	3.20E-04
17. Gun	0.862	8.34E-01
21. Tattoos	0.821	9.02E-01
15. Attached object category	0.721	9.84E-01
18. Object in hand	0.721	9.84E-01
3. Face covered	N/A	N/A
16. Bag	N/A	N/A
19. Gloves	N/A	N/A

(b) Comparative traits

Clothing attribute	**F-ratio** (df = 443)	**p-value** ($p \leq 0.05$)
2. Head coverage	15.670	3.68E-86
8. Sleeve length	12.021	1.02E-71
11. Leg length	4.819	2.49E-30
16. Bag	4.689	2.42E-29
3. Face covered	2.559	9.38E-12
7. Neckline size	2.515	2.26E-11
14. Heel level	1.636	2.73E-04

Table 3.2: MANOVA results of soft clothing traits showing significance by different standard statistics

MANOVA statistic	**(a) Categorical traits**		**(b) Comparative traits**	
	F-ratio (df = 443)	**p-value** ($p \leq 0.05$)	**F-ratio** (df = 443)	**p-value** ($p \leq 0.05$)
Pillai's Trace	2.314	<0.0001	3.343	<0.0001
Wilks' Lambda	3.314	<0.0001	4.729	<0.0001
Hotelling's Trace	5.186	<0.0001	6.853	<0.0001
Roy's Largest Root	23.519	<0.0001	20.045	<0.0001

3.4 Mutual Dependence Analysis of Clothing Traits

3.4.1 Statistical Dependency

Statistical Dependency (SD) is a method used to examine whether the values of a feature are dependent on the associated class labels and whether those feature values and class labels are unlikely to co-occur by chance [94]. Here in our biometric context, the feature is a soft clothing trait and the class labels represent the subject IDs. The larger the SD value, the higher is the dependency between trait values and subject IDs. While when a trait is fully independent of the subject IDs, it receives a minimal value of SD ≤ 1 and reflects insignificant quantity of dependence. The SD measure is calculated for each soft clothing trait and can be defined as:

$$\mathrm{SD} = \sum_{x \in X} \sum_{y \in Y} p(x, y) \frac{p(x, y)}{p(x)p(y)} \tag{3.3}$$

where X a variable denoting a trait value, and Y a variable representing the class label (i.e. the subject's identity). $p(x, y)$ is the joint probability density function of the random variables X and Y, where $p(x)$ and $p(y)$ are the marginal probability density functions respectively.

Table 3.3: Statistical dependency (SD) analysis of soft clothing traits

(a) Categorical traits	
Clothing attribute	**SD**
5. Upper body clothing category	2.483
6. Neckline shape	1.949
20. Style category	1.946
8. Sleeve length	1.854
10. Shape (of lower clothing)	1.765
7. Neckline size	1.732
11. Leg length (of lower clothing)	1.504
13. Shoes category	1.409
12. Belt presence	1.323
1. Head clothing category	1.195
2. Head coverage	1.195
9. Lower body clothing category	1.094
14. Heel level	0.926
4. Hat	0.861
21. Tattoos	0.757
15. Attached object category	0.350
18. Object in hand	0.350
17. Gun	0.018
3. Face covered	1.29E-04
16. Bag	1.29E-04
19. Gloves	1.29E-04

(b) Comparative traits	
Clothing attribute	**SD**
8. Sleeve length	4.524
2. Head coverage	4.310
16. Bag (size)	4.125
3. Face covered	3.927
7. Neckline size	3.813
14. Heel level	3.765
11. Leg length	3.463

Table 3.3 shows categorical and comparative soft clothing traits ordered by the SD measure. It appears natural that the comparative traits are more statistically dependent by receiving higher SD values than the categorical traits. Upper body clothing category (5) appears as the best categorical trait in terms of SD with a far high score, followed next by neckline shape (6) and style category (20) both with almost the same score. Sleeve length (8) is the highest dependent comparative trait and among the highest categorical traits, which may emphasize the usefulness of this trait.

3.4.2 Mutual Information

Mutual information (MI) is a method used to measure the mutual dependence of two variables. Note that the concept and the way of calculation of MI is very similar to the statistical dependency (SD) explained in Section 3.4.1. It is worth noting that analysing mutual dependence among soft clothing attributes via SD or MI has not been applied prior to this research. SD may be deemed more sensitive and preferable in some high-dimensional classification problem like voice analysis [94]. Nevertheless, MI is more common in use and meant to provide more accurate translation of mutual dependence between variables by enforcing the logarithmic compression shown in Eqn. (3.4).

$$\mathrm{MI} = I(X,Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)log\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{3.4}$$

The MI quantity (usually denoted as $I(\mathrm{X},\mathrm{Y})$) is computed here in two aspects: (i) similar to the way used to compute SD (Section 3.4.1), considering X as a variable denoting a trait value, and Y a variable representing the subject's identity; and (ii) considering both X and Y as trait values of two different soft clothing traits, resulting in measuring MI of these two traits. $p(x,y)$ is the joint probability density function of the random variables X and Y, where $p(x)$ and $p(y)$ are the marginal probability respectively. Table 3.4 shows categorical and comparative soft clothing traits ordered by MI deduced (based on aspect (i)) for each trait against the subject IDs. While Fig. 3.5 shows a symmetric colour-coded matrix of mutual information between each two soft clothing traits (based on aspect (ii)).

It can be observed that analysis of MI in Table 3.4 presents different ranking of clothing traits compared with the trait ranking presented by SD in Table 3.3, suggesting that MI and SD differently measure the mutual dependence, even though they are based on the same concept. Upper body clothing category (5), neckline shape (6), and sleeve length (8) remain on top of the rank unlike style category (20) which fall back in rank. Categorical leg length (11) interestingly becomes the next high mutually informative trait after sleeve length (8), whereas it is the lowest among the comparative traits.

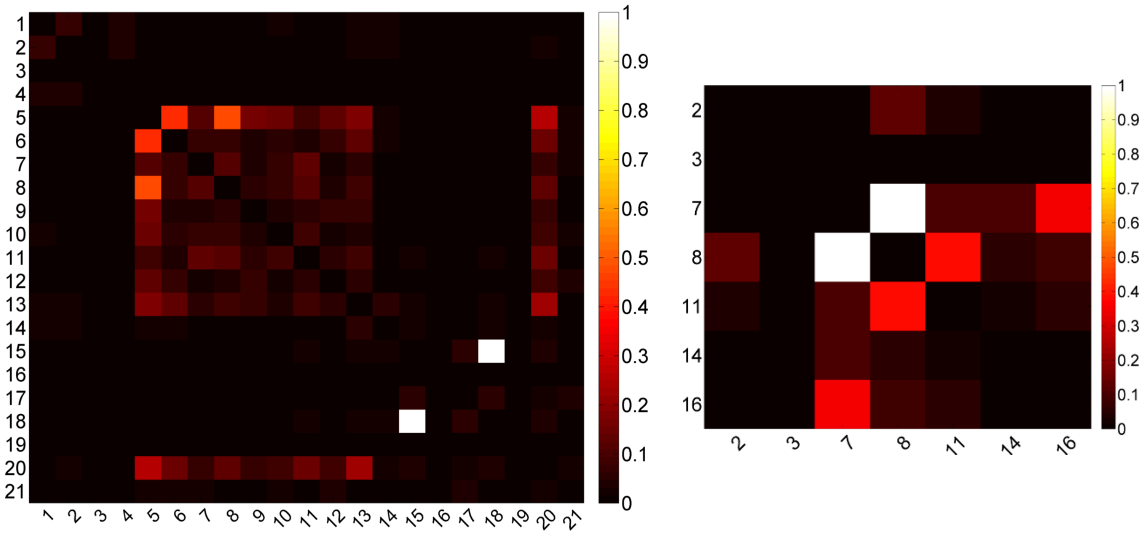
Table 3.4: Mutual information (MI) analysis of soft clothing traits

(a) Categorical traits

Clothing attribute	**MI** $I(X,Y)$
5. Upper body clothing category	1.387
6. Neckline shape	1.289
8. Sleeve length	1.130
11. Leg length (of lower clothing)	0.993
13. Shoes category	0.976
20. Style category	0.972
10. Shape (of lower clothing)	0.946
7. Neckline size	0.843
12. Belt presence	0.815
9. Lower body clothing category	0.329
14. Heel level	0.293
21. Tattoos	0.246
15. Attached object category	0.202
18. Object in hand	0.202
1. Head clothing category	0.067
2. Head coverage	0.067
4. Hat	0.045
17. Gun	0.001
3. Face covered	1.44E-06
16. Bag	1.44E-06
19. Gloves	1.44E-06

(b) Comparative traits

Clothing attribute	**MI** $I(X,Y)$
8. Sleeve length	2.288
2. Head coverage	2.242
16. Bag (size)	2.184
3. Face covered	2.116
14. Heel level	2.076
7. Neckline size	2.042
11. Leg length	1.836



(a) Categorical clothing traits (b) Comparative clothing traits

Figure 3.5: Mutual information (MI) matrix of soft clothing traits

In Fig. 3.5, normalized MI measurements (ranging from 0 to 1) of all possible pairs of soft clothing traits are visually illustrated, where bright colours (white, yellow) reflect high MI values, vivid colours (orange, red) display medium values, and dark colours (brown, black) represent low values. The diagonal quantities are all set to zero (black), enforcing the notion that a variable cannot reduce the uncertainty of itself. As such, in the categorical matrix, it can be perceived that the traits (15) and (18) relating to the description of objects attached to the body are the highest mutually informative traits. Most observable MI values are located between nine traits (5-13), where each provides some mutual information with the other eight, ranging from low to medium. Upper clothing category (5) shows well mutual information with neckline shape (6) and sleeve length (8), signifying the compatibility between a certain upper clothing type and the specification of its parts (neckline, sleeve etc.). Clothing style category (20) as a general description offers some notable mutual information with upper and lower body clothing traits (from 5 to 13). This intuitively reflects the reality that detailed characteristics of upper and lower clothing explicitly contribute to inform what clothing style is worn by a person, and the way around, a particular clothing style can give an idea about likely upper and lower clothing descriptions. On the other hand, in the comparative matrix, the highest MI quantity is between neckline size (7) and sleeve length (8). The next significant mutual information exists between neckline (7) and bag (16) sizes, and sleeve (8) and leg (11) lengths.

3.5 Analysis of User-annotators' Labelling Behaviour

Analysing annotator performance is necessary to study the accuracy of human-based labelling data. Such analysis can provide insight into the reliability of collected data from different perspectives [95, 96]. It is worth analysing the effects from labeller intervention. A framework for analysis can be designed and applied to annotators participating in a relatively similar task such as word sense annotation in Natural Language Processing (NLP) [97]. This allows investigation of annotator behaviour leading to: (i) determination of outliers; (ii) to cluster annotators by behaviour; and (iii) further to recognise confusable labels [97]. Further analysis may even lead to investigation of the semantic gap between annotators, caused by reasons including gender and age differences [98]. Moreover, the importance and benefit of annotator analysis is emphasised with crowd labelling data collected from untrained Mechanical Turk labellers, leading to better understanding and improvement of annotation data [99]. While it has been shown that trained annotators are superior to a larger number of untrained annotator or Mechanical Turk labellers, in providing more reliable labelling [97].

Compared with word sense annotation NLP mechanisms [96, 97, 100], the structure of our clothing annotation mechanism is likely to have smaller room of subjectivity and systematic differences among a subset of annotators. That is due to two major characteristics. Firstly, a limited set of labels in multiple-choice wise is provided for an annotator to describe an attribute. Secondly, the labelling task exclusively depends on a single displayed subject or two displayed subjects (in the comparison task), as an annotator is asked to exactly describe what they see not what they infer, which tends to minimise the variation of judgement affected by their natural differences in opinion and behaviour.

Clothing labels and comparisons were collected from 27 user-annotators (19 males and eight females). With regard to the acquired comparison data, we are not yet able to perform analysis of user-annotators' labelling behaviour through the available comparative data, since we asked each user-annotator to compare ten pairs of subjects differ from those compared by the remaining user-annotators, aiming to cover as much different comparisons as we can form all possible pairs. On the other hand, in the acquired (categorical) label data, each user-annotator was asked to annotate only a subset of ten randomly selected subjects in database. Therefore, there is no available annotation for all 128 subjects provided by each of the 27 user-annotators. Due to this limitation, we are not yet able to provide a comprehensive analysis for all user-annotators' labelling behaviour through the available categorical data. As such, only partial analysis of user-annotators' behaviour is possible. We selected two groups of user-annotators *M*, who provided categorical labels for the same *N* group of subjects. Thereby, two subsets of annotation data describing a number of subjects, where each subject has multiple annotations provided by exactly the same group of user-annotators. These groups are detailed in Table 3.5. This analysis has not yet been applied in the existing research to human-based clothing annotations for either biometric or non-biometric purposes.

Table 3.5: Selected group of user-annotators labelling the same set of subjects

Group		**User-annotators**	**Subjects**	**Number of labels**	
ID	**Description**			***Cat-21***	***Cat-7***
Group-1	3 annotators X 10 subjects	U03, U05, U26	051, 052, 053, 054, 055, 056, 057, 058, 059, 060	630	210
Group-2	4 annotators X 5 subjects	U05, U11, U21, U22	091, 092, 093, 094, 095	420	140

In order to analyse inter-annotator agreement in the context of soft clothing attribute annotation, we use a statistical (consensus) measurement of Cohen's kappa [101], which is suitable for categorical data [102], and was previously used in the context of NLP annotation tasks [103, 104]. While kappa agreement measurement is deemed to reflect more robust score than simple calculated percentage of agreement. Table 3.6 shows per attribute both percent agreement and kappa agreement of the two groups of user-annotators separately, where attributes are ordered by kappa value. Kappa agreement (κ) score can be calculated as follows:

$$\kappa = \frac{P(a) - P(e)}{1 - P(e)} \tag{3.5}$$

where $P(a)$ represents the proportion of observed agreement among annotators. $P(e)$ is the proportion of agreement on a certain label by chance, which is equivalent to the probability of selecting a dominant agreed label from a set of N available labels, so it can be calculated as $1/N$ because each label has an equal probability (chance) among a set of designated labels per attribute. In Table 3.6, it can be observed that the inter-annotator agreement varies in attribute-level ranging from 0.5 to 1.0. While the received total average scores of kappa agreement (0.82 in Group-1) and (0.79 in Group-2) indicate a high overall consensus among inter-annotator labelling. Fig. 3.6 compares Kappa (κ) agreement scores between Group-1 and Group-2, where clothing attributes are ordered by average score of both groups.

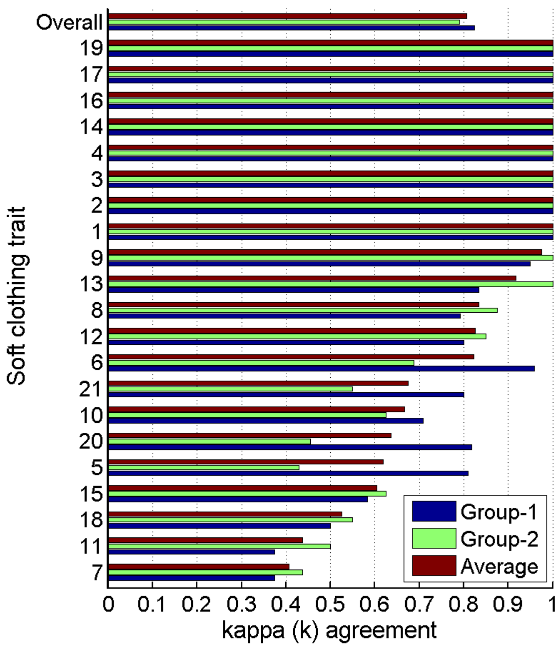


Figure 3.6: Kappa (κ) agreement scores of Group-1 and Group-2 per attribute ordered by average score

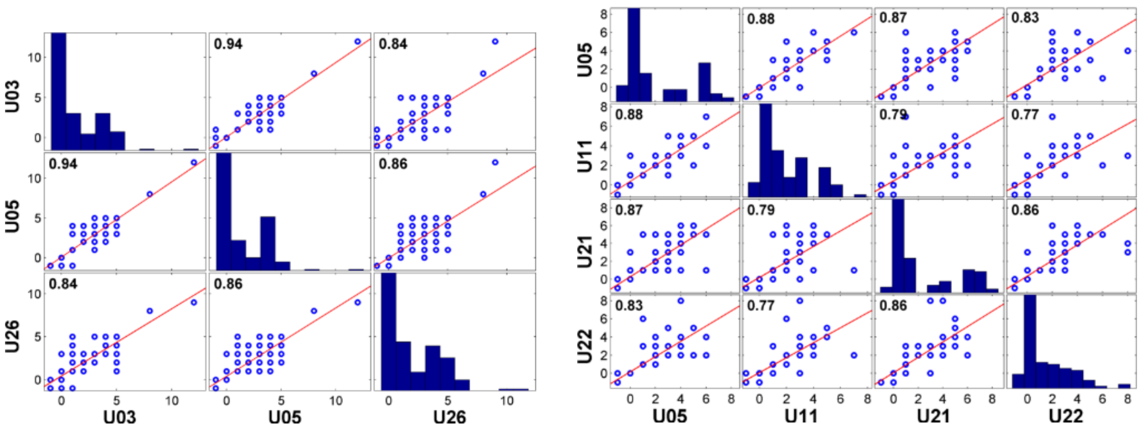
Table 3.6: Inter-annotator agreement on clothing labelling ordered by Kappa (κ) agreement

Group-1

Clothing attribute	Labels: Available	Labels: Used	Agreement: (%)	Agreement: Kappa (κ)
1. Head clothing category	5	1	1.00	1.00
2. Head coverage	5	1	1.00	1.00
3. Face covered	3	1	1.00	1.00
4. Hat	3	1	1.00	1.00
14. Heel level	4	1	1.00	1.00
16. Bag	6	1	1.00	1.00
17. Gun	3	1	1.00	1.00
19. Gloves	3	1	1.00	1.00
6. Neckline shape	5	3	0.97	0.96
9. Lower body clothing category	3	2	0.97	0.95
13. Shoes category	5	4	0.87	0.83
20. Style category	12	6	0.83	0.82
5. Upper body clothing category	8	5	0.83	0.81
12. Belt presence	3	3	0.87	0.8
21. Tattoos	3	2	0.87	0.8
8. Sleeve length	5	4	0.83	0.79
10. Shape (lower clothing)	5	4	0.77	0.71
15. Attached object category	5	2	0.67	0.58
18. Object in hand	3	2	0.67	0.50
7. Neckline size	5	4	0.50	0.38
11. Leg length (lower clothing)	5	5	0.50	0.38
Overall	**99**	**54**	**0.86**	**0.82**

Group-2

Clothing attribute	Labels: Available	Labels: Used	Agreement: (%)	Agreement: Kappa (κ)
1. Head clothing category	5	1	1.00	1.00
2. Head coverage	5	1	1.00	1.00
3. Face covered	3	1	1.00	1.00
4. Hat	3	1	1.00	1.00
9. Lower body clothing category	3	1	1.00	1.00
13. Shoes category	5	3	1.00	1.00
14. Heel level	4	1	1.00	1.00
16. Bag (size)	6	1	1.00	1.00
17. Gun	3	1	1.00	1.00
19. Gloves	3	1	1.00	1.00
8. Sleeve length	5	3	0.90	0.88
12. Belt presence	3	3	0.90	0.85
6. Neckline shape	5	3	0.75	0.69
10. Shape (lower clothing)	5	4	0.70	0.63
15. Attached object category	5	2	0.70	0.63
18. Object in hand	3	2	0.70	0.55
21. Tattoos	3	2	0.70	0.55
11. Leg length (lower clothing)	5	2	0.60	0.50
20. Style category	12	6	0.50	0.45
7. Neckline size	5	4	0.55	0.44
5. Upper body clothing category	8	8	0.50	0.43
Overall	**99**	**51**	**0.83**	**0.79**



(a) Correlation matrix of **Group-1** (b) Correlation matrix of **Group-2**

Figure 3.7: Correlations between user-annotators' labelling behaviour

We also use label data of Group-1 and Group-2 to explore the correlation between annotators not between clothing attributes. This provides an additional indicator of the reliability and consistency of inter-annotator data describing the same subject. Fig. 3.7 shows scattered correlations between annotators of each group and the histogram of labels provided by all annotators describing the same subjects of that group. It is obvious that the labelling behaviour in both groups are highly and positively correlated between annotators in the same group ranging from 0.77 to 0.94. In sum, with respect to the current sample analysis results, we found well correlated inter-annotator labels and high inter-annotator agreement, which suggest that the clothing annotation data has reasonable consistency and good reliability.

3.6 Conclusions

In this chapter, extended analysis from different perspectives are applied on soft clothing traits. The trait significance and correlations between clothing labels and derived soft biometrics are investigated. In general, we find that the relative attributes underlying all the proposed clothing-based approaches carry the most significant and effective information to describe clothing. A good correlation between two attributes allows for the prediction of a missing attribute of those most likely from the other.

Clothing trait data distributions are described and discussed from statistical point of view. ANOVA test takes place in this analytical study as a major elicitation of the discrimination capabilities of soft clothing traits. The statistical dependency and mutual information are explored and the most useful and informative traits are highlighted and compared. Furthermore, user annotators' labelling behaviour is studied and analysed in terms of inter-annotator correlations and consensus, signifying insightful information about how understandable and usable are the proposed semantic clothing attributes, and how accurate and reliable is the collected human-based clothing label data.

Chapter 4

Identification using Soft Clothing Traits

4.1 Augmenting Soft Biometrics with Clothing

In this experiment, we employ semantic clothing attributes as soft biometric traits for subject identification and explore the validity and efficiency of clothing descriptions for identification. We use our three proposed approaches (*Cat-21*, *Cat-7* and *Cmp*) of soft clothing traits to enhance the identification performance of soft body traits (described in Section 2.2.1) in two different aspects. The first aspect is to augment only the traditional soft traits (Age, Ethnicity, Sex, and Skin Colour) which were the four traits grouped in the *Global* category in [39]. The second aspect is to augment all the aforementioned 23 soft body traits including the four traditional traits.

For each subject we form a single 23-trait feature vector from the multiple annotations of that subject to build a gallery of soft body descriptions for all subjects (referred to as *softBody*). From *softBody* we derive another gallery of a traditional four-trait feature vector per subject (referred to as *tradSoft*) comprising only traditional soft descriptions.

Both *tradSoft* and *softBody* are extended to three other versions supplemented by the proposed soft clothing traits. Each feature vector describing a single subject in *tradSoft* is concatenated to a corresponding feature vector describing the same subject in each of *Cat-21*, *Cat-7* and *Cmp* to compose three versions of augmented galleries: *tradCat-21*, *tradCat-7* and *tradCmp* respectively. The same process is applied to *softBody*, resulting in its three augmented versions *softCat-21*, *softCat-7* and *softCmp*. Each *softBody* and *tradSoft* is tested for identification separately and then compared to its three augmented versions. For testing, proper probe sets are prepared for *softBody* and *tradSoft*, so that for each subject we select a number of annotations from body-descriptions data equal to the number of annotations collected for clothing descriptions. These probe sets are extended to their corresponding augmented versions, according to their galleries, by concatenating them to each of the three clothing probe sets to enable the testing of their augmented galleries. Based on the leave-one-out method, human identification using soft body approaches and their augmented counterparts is achieved and evaluated. Table 4.1 lists and describes the used soft biometric approaches (and produced galleries) of body traits and when combined with clothing traits.

Table 4.1: Body and clothing approaches used for testing and comparing identification performance

Clothing-based biometrics	Description
Cat-21	21 categorical soft clothing biometrics (binary/multiclass, and relative)
Cat-7	7 categorical soft clothing biometrics (relative)
Cmp	7 comparative soft clothing biometrics (relative)
Body-based biometrics	**Description**
tradSoft	4 categorical soft body biometrics (Age, Ethnicity, Sex, and Skin Colour)
softBody	23 categorical soft body biometrics including ***tradSoft***
Combined clothing and body biometrics	**Description**
tradCat-21	***Cat-21*** combined with ***tradSoft***
softCat-21	***Cat-21*** combined with ***softBody***
tradCat-7	***Cat-7*** combined with ***tradSoft***
softCat-7	***Cat-7*** combined with ***softBody***
tradCmp	***Cmp*** combined with ***tradSoft***
softCmp	***Cmp*** combined with ***softBody***

Leave-one-out testing is used to evaluate the performance of clothing traits. Thereby, user annotations are used in turn as a query to identify and retrieve matching subjects from a tested gallery. Hence, two *Cat-N* based probe sets are designed such that, each subject's annotation provided by a single annotator, is normalized and formulated as 21-label and 7-label feature vectors to be used to probe *Cat-21* and *Cat-7* galleries. The same 7-label probe set used for testing *Cat-7* is used to probe the *Cmp* gallery, except that each feature vector in the probe set is mapped to its comparative-traits vector predicted by *w*. This mapping is performed to produce the third *Cmp*-based probe set containing a new form of testing feature vector that is able to probe the *Cmp* gallery.

4.1.1 Performance Evaluation

The likelihood between every single probe-vector and all subject-vectors in a gallery is estimated and retrieved, resulting in an ordered list of all subjects based on the likelihood evaluated by the sum of Euclidean distance between probe and gallery vectors, which can be defined as $d(p,g) = \sqrt{\sum_{i=1}^{n}(p_i - g_i)^2}$, where $d(p,g)$ is the distance between the *probe* (or query) feature vector p and the *gallery* feature vector g, where both p and g vectors have the same number (and order) of n features. As such, the resulting distance $d(p, g)$ is a value representing the sum of all n differences, computed between each i^{th} feature of p and g vectors. The smaller the distance value of $d(p, g)$, the more is the likelihood between the two vectors (or subjects) p and g. In all forthcoming experiments, we use one or more query feature-vectors (as testing samples) per subject in each probe set, while we use only a single feature vector (as an enrolled soft biometric signature) for each subject per gallery.

It is noteworthy that the standard and most commonly use of Euclidean distance matching is when the variables are true measurements usually represented with continuous values belonging to the set of real number $\mathbb{R}$ for which the calculation of Euclidean distance makes most sense, and it makes reasonably good sense if variables are ranks like in our context when we match two vectors of comparative traits, which are represented by real numbers known as relative measurements. However, Euclidean distance matching is still usable as a measure of similarity/dissimilarity with most kinds of variables including presence/absence or binary (two-category) variables, such as '*Belt presence*', can be treated meaningfully in the calculation of Euclidean distance.

In our context, we have two kinds of categorical traits: relative and non-relative. For a relative trait such as (*Sleeve length*: 'Very short', 'Short', 'Medium', 'Long', 'Very long'), we assign a set of ordered values ranging from 1 to 5 corresponding to ordered categorical labels to reflect the degree-of-strength of that trait. In this case, even though they are not true measurements or ranks, the calculation of Euclidean distance still makes reasonable sense due to the consistent and logical coding for the (ordinal) labels as a bipolar scale. On the other hand, for a (multi-class) non-relative trait such as (*Upper body clothing category*: 'Jacket', 'Jumper', 'T-shirt' etc.), we assign a set of integer values starting from 1 up to the number of labels to numerically represent textual expressions of the labels. In this case, the Euclidean distance matching is not used in its ideal feature space, since the assigned values of such pure (nominal) labels do not reflect any ordering or measurements; rather they can be assigned values in any order or their assigned values can be exchanged. Here, we consistently assign label values whenever they are used to describe subjects, besides we perform the likelihood matching between full feature vectors (or biometric signatures) including a number of non-relative and relative traits as well. As such, we can reasonably and appropriately use the calculation of Euclidean distance to achieve desired matching and successful recognition.

The Cumulative Match Characteristic (CMC) curve is applied to summarize the identification accuracy, via the k Nearest Neighbour (kNN) approach based on the likelihood evaluation described earlier, which scores the existence of the correct subjects within a likelihood-based ranked list (from the nearest neighbour to the furthest neighbour) starting from list length of k = 1 to k = 128 the total number of subjects in a tested gallery. Receiver Operator Characteristic (ROC) analysis is used to assess and compare the approaches performance and their generated errors. A set of further performance metrics are deduced from the ROC analysis comprising the Area under the Curve (AUC) as in our consideration a smaller are under the ROC curve reflecting a less error and a better performance, and the Equal Error Rate (EER). Also the Decidability Index (d') metric is computed from the normalized distance between the two means of *Genuine* (G) and *Imposter* (I) distributions such that $d' = |\mu_G - \mu_I| / \sqrt{(\sigma_G^2 + \sigma_I^2)/2}$. Eventually, with respect to all metrics, all approaches are ranked by their overall performance. According to the way we currently use to infer these performance metrics, it is worth noting that all CMC scores and d' are desired to be large to indicate better performance, whereas the smaller the values of EER, AUC, and (the rank of achieving 100% accuracy), the higher is the performance.

Towards further understanding of identification performance, additional analysis is provided for genuine versus imposter and their caused errors. The first analysis is a representation for estimated distributions of matching scores in terms of intra-class (genuine) and inter-class (imposter). The second analysis demonstrates correspondingly the percentage of False Acceptance Rate (FAR) versus the percentage of False Rejection Rate (FRR) under multiple thresholds ranging from 0 to 1. The low intra-class variance of biometric traits indicates that the used traits seem to maintain permanence and repeatability, in contrast the high inter-class variance indicates that the used traits tend to successfully distinguish subjects [105]. Note that this performance evaluation method is also used throughout all identification/retrieval experiments in this thesis.

4.1.2 Augmenting Traditional Soft Biometrics

Table 4.2 reports the CMC match scores and average-sum scores in different ranks for the traditional soft traits (Age, Ethnicity, Sex, and Skin Colour) and when augmented by clothing. Fig. 4.1 shows the CMC performance of each approach. Adding clothing to traditional soft biometrics consistently enhances identification with up to 59% offered by the categorical traits *tradCat-21*, which achieves the highest scores in all ranks. *tradCat-7* comes next in performance, then *tradCmp* with slightly lower scores. Fig. 4.2 shows ROC curves of the augmenting approaches have better accuracy and less error, where *tradCat-21* consistently presents the highest accuracy, followed by *tradCat-7*. While Table 4.2 indicates the smallest Equal Error Rate (EER) and the Area Under the ROC Curve (AUC) for *tradCat-21*, *tradCmp* receives the best score in the decidability metric *d'*; all the best values are shown in bold. Fig. 4.3 presents genuine and imposter distributions along with their consequent acceptance and rejection error rates, which clarifies the matching performance of clothing approaches.

Table 4.2: CMC scores and ROC metrics of traditional soft biometrics and when augmented by clothing

Approach	**Top rank**	**AVG sum match scores up to rank**		**100% accuracy achieved at rank**	**EER**	**AUC**	***d'***	**Overall rank**
	=1	***=10***	***=128***					
tradSoft	0.16	0.31	0.866	72	0.181	0.136	1.734	4
tradCat-21	**0.75**	**0.92**	**0.993**	**29**	**0.097**	**0.036**	1.789	**1**
tradCat-7	0.47	0.76	0.974	37	0.108	0.050	2.010	2
tradCmp	0.44	0.74	0.966	63	0.161	0.069	**2.045**	3

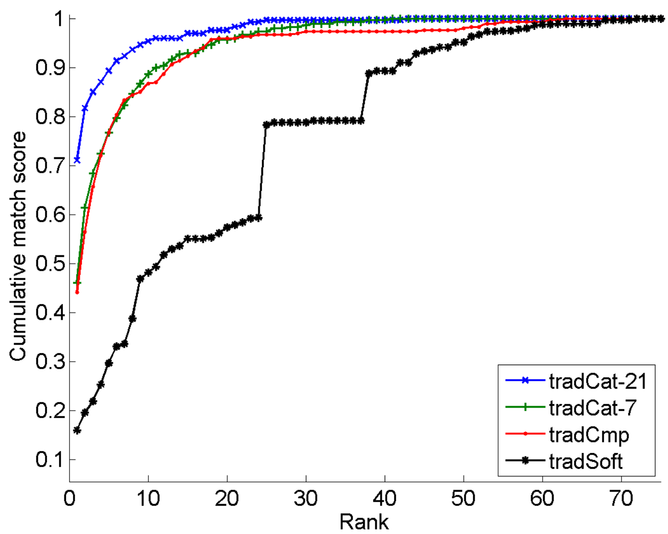


Figure 4.1: CMC (up to rank 75) of traditional soft biometrics and when augmented by clothing

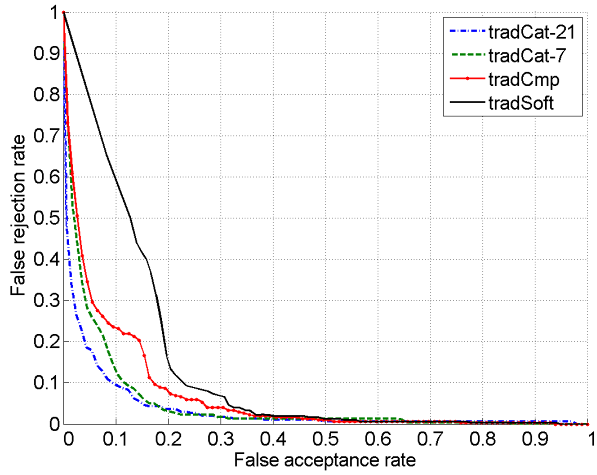
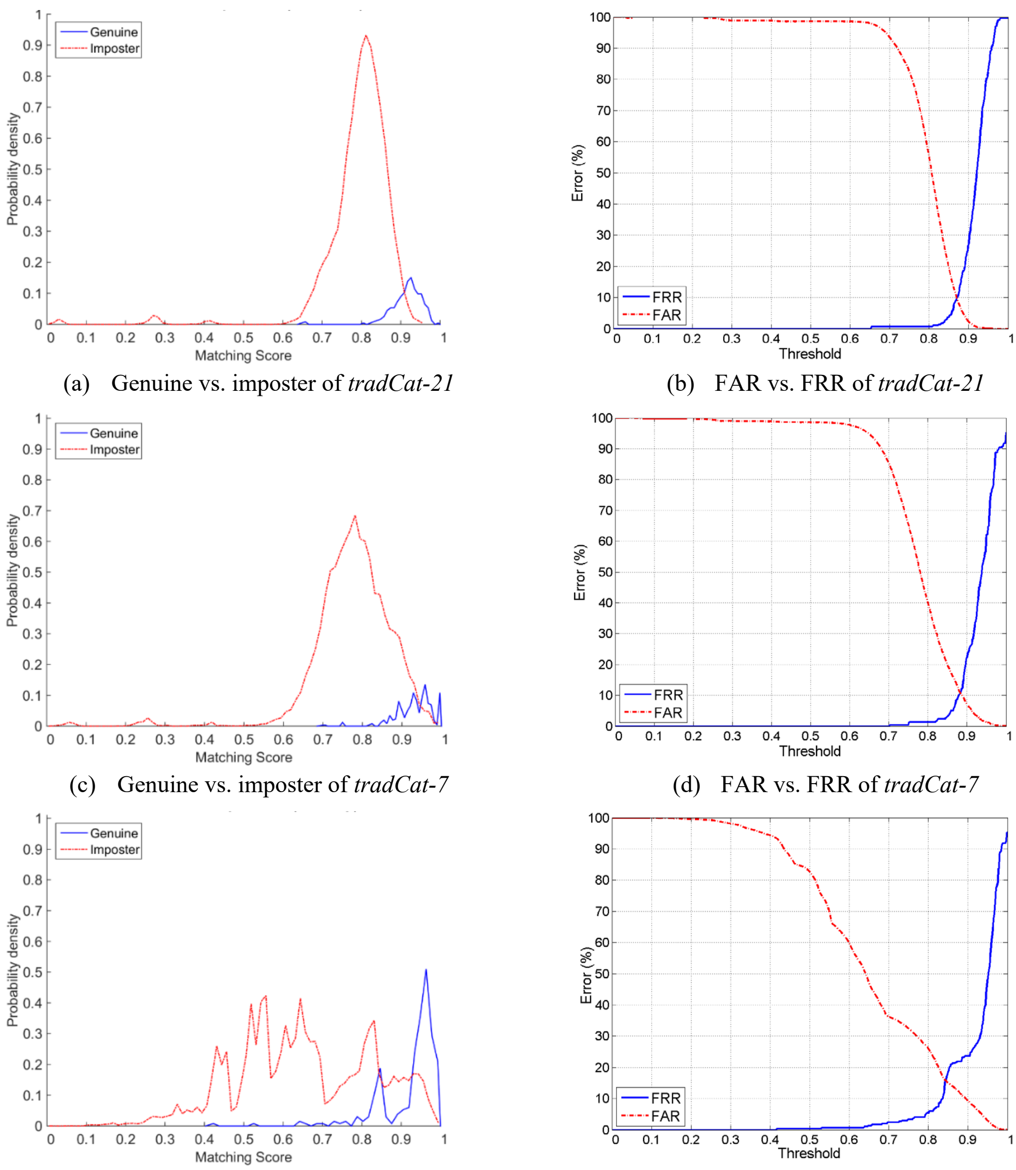


Figure 4.2: ROC performance of traditional soft biometrics and when augmented by clothing



(a) Genuine vs. imposter of *tradCat-21*

(b) FAR vs. FRR of *tradCat-21*

(c) Genuine vs. imposter of *tradCat-7*

(d) FAR vs. FRR of *tradCat-7*

(e) Genuine vs. imposter of *tradCmp*

(f) FAR vs. FRR of *tradCmp*

Figure 4.3: Genuine vs imposter and error curves of clothing traits added to *tradSoft* for identification

4.1.3 Augmenting Soft body Biometrics

The CMC curves in Fig. 4.4 and the results in Table 4.3 represent the identification performance using soft body traits with performance augmented by soft clothing traits. The performance of the soft body traits is considerably improved by adding the clothing yielding an increase ranging from 7% to 17% in the rank 1 result. The identification rate of *softBody* jumps from 78% to 95% when augmented by categorical traits *softCat-21,* and reaches 100% rapidly at rank 9. Likewise, augmenting the traditional soft approach, *softCat-7* remains, achieving the second best performance and the comparative measures *softCmp* holds the highest decidability *d'*. Fig. 4.5 illustrates the resulting ROC performance of each approach. For each clothing approach, Fig. 4.6 shows genuine and imposter densities besides FAR and FRR curves.

Table 4.3: CMC scores and ROC metrics of the soft body biometrics and when augmented by clothing

Approach	**Top rank**	**AVG sum match scores up to rank**		**100% accuracy achieved at rank**	**EER**	**AUC**	***d'***	**Overall rank**
	=1	***=10***	***=128***					
softBody	0.78	0.92	0.991	37	0.087	0.028	2.785	4
softCat-21	**0.95**	**0.99**	**0.999**	**9**	**0.050**	**0.014**	2.634	**1**
softCat-7	0.88	0.96	0.996	32	0.063	0.018	2.814	2
softCmp	0.85	0.94	0.994	36	0.080	0.026	**2.827**	3

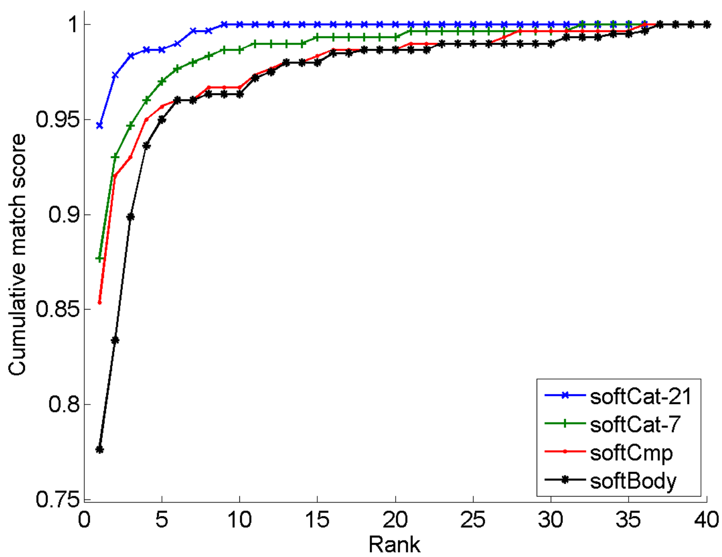


Figure 4.4: CMC performance (up to rank 40) of soft body biometrics and when augmented by clothing

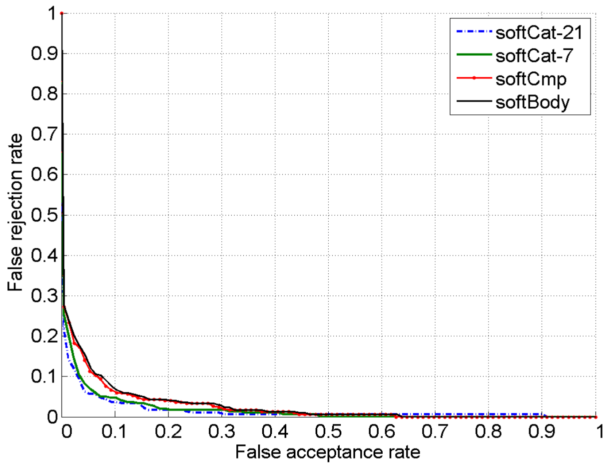
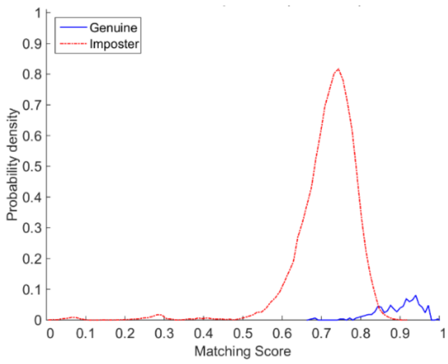
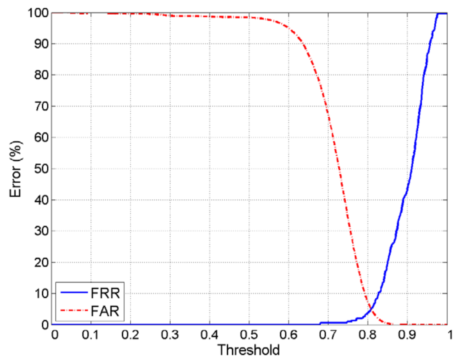


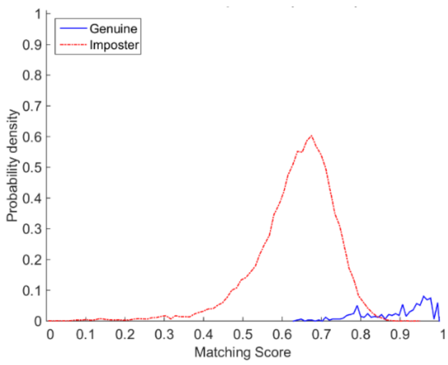
Figure 4.5: ROC performance of soft body biometrics and when augmented by clothing



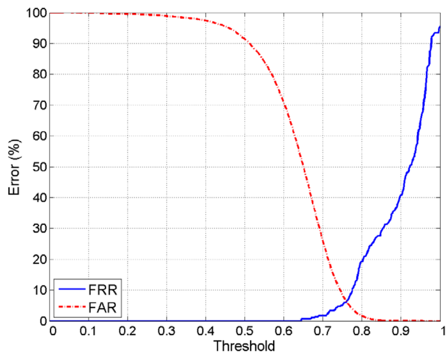
(a) Genuine vs. imposter of *softCat-21*



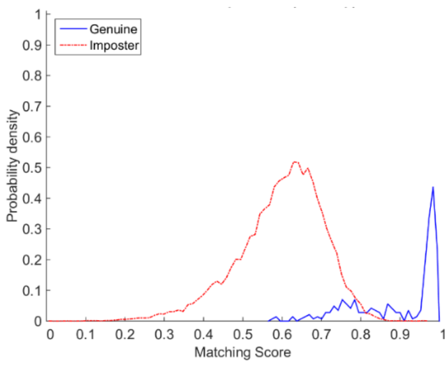
(b) FAR vs. FRR of *softCat-21*



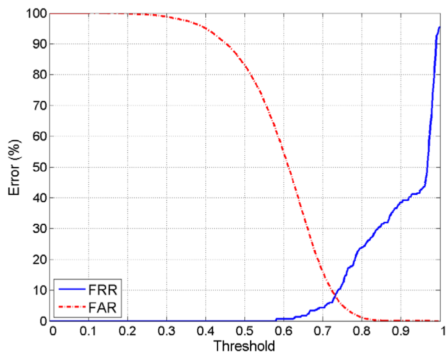
(c) Genuine vs. imposter of *softCat-7*



(d) FAR vs. FRR of *softCat-7*



(e) Genuine vs. imposter of *softCmp*



(f) FAR vs. FRR of *softCmp*

Figure 4.6: Genuine vs imposter and error curves of clothing traits added to *softBody* for identification

4.2 Identification using Soft Clothing Traits Alone

Each of the proposed approaches (*Cat-21*, *Cat-7* and *Cmp*) is used in isolation for human identification, when soft clothing traits are the only used biometric. Here the leave-one-out method is also applied to probe the three soft clothing galleries using the appropriate probe set for each. The CMC scores and the ROC analysis of the proposed approaches are given in Table 4.4. Although all the approaches tend to improve the identification score sharply throughout the rank increase from 1 to 10, the categorical labels *Cat-21* start from a much better score and gain a much higher average score than *Cat-7* and *Cmp* as shown in Fig. 4.8 comparing their CMC performance. The *Cat-21* approach outperforms the other approaches in all terms, but the decidability d' is the largest and best in *Cmp*. It is noteworthy that unlike in augmenting soft biometrics, the comparative traits *Cmp* achieve a better performance than their categorical counterparts underlying *Cat-7*. Also in terms of ROC performance, Fig. 4.9 illustrates that *Cmp* accordingly received less error than *Cat-7*.

(a) Query subject, recognised at rank 1

(b) Left: Query subject, recognized at rank 65.
Right: confused closest match at Rank 1

Figure 4.7: Matching performance by clothing alone: (a) good match, and (b) poor match

Fig. 4.7 shows a good and poor samples of identification by clothing, where (a) a subject successfully recognised by given query annotations at rank1, whereas (b) by given query annotations describing the left subject, he is poorly recognized at rank 65, while the right subject was incorrectly the closest match. Fig. 4.10 shows genuine and imposter distributions and highlights the trade-offs between FAR and FRR curves of each approach, which confirms the potency of *Cat-21* traits in identification either in fusion or alone, as shown also in Fig. 4.3 and 4.6. The intra-class variance of *Cat-21* is lower and better than *Cat-7* and *Cmp*, whereas the inter-class variance of *Cmp* is the highest.

Accordingly, *Cat-21* provides the smallest EER in identification. Here, *Cmp* compared with *Cat-7* achieves slightly better identification and receives relatively smaller EER. That is obviously consistent with numerical scores in Table 4.4.

Table 4.4: CMC scores and ROC metrics of soft clothing biometrics when used alone for identification

Approach	**Top rank**	**AVG sum match scores up to rank**		**100% accuracy achieved at rank**	**EER**	**AUC**	***d'***	**Overall rank**
	=1	***=10***	***=128***					
Cat-21	**0.63**	**0.843**	**0.984**	**41**	**0.137**	**0.059**	1.442	**1**
Cat-7	0.27	0.507	0.923	92	0.192	0.108	1.303	3
Cmp	0.28	0.510	0.929	96	0.174	0.088	**1.824**	2

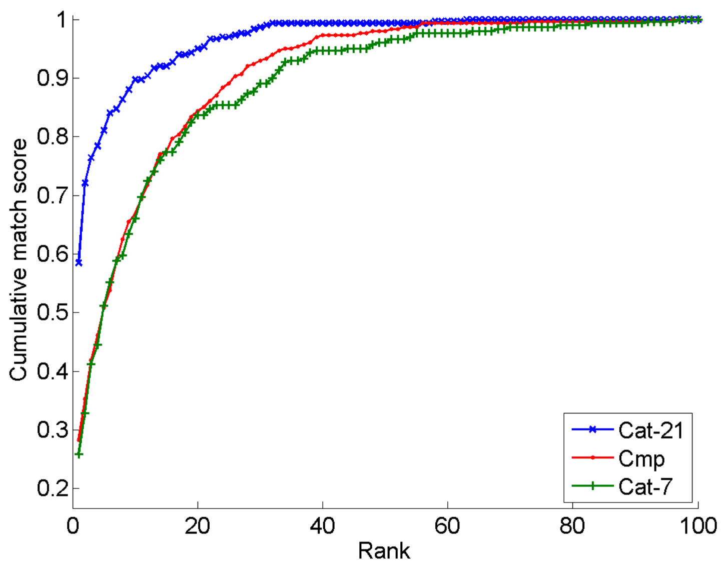


Figure 4.8: CMC performance (up to rank 100) of soft clothing biometrics when used alone

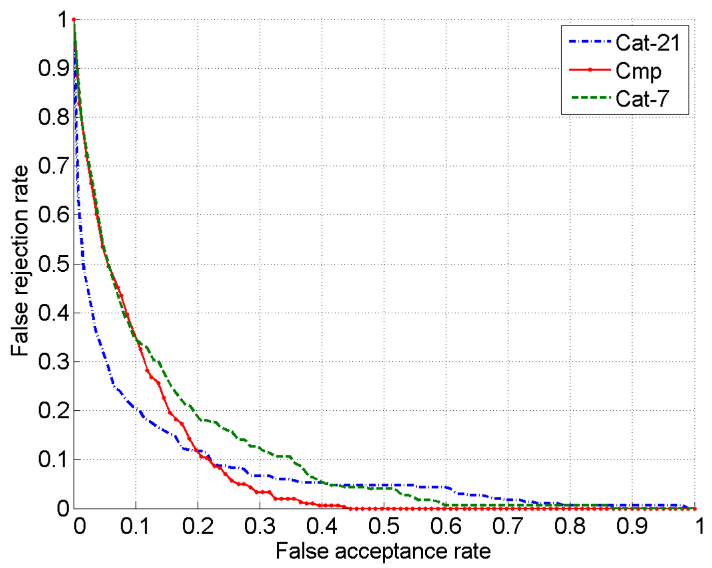
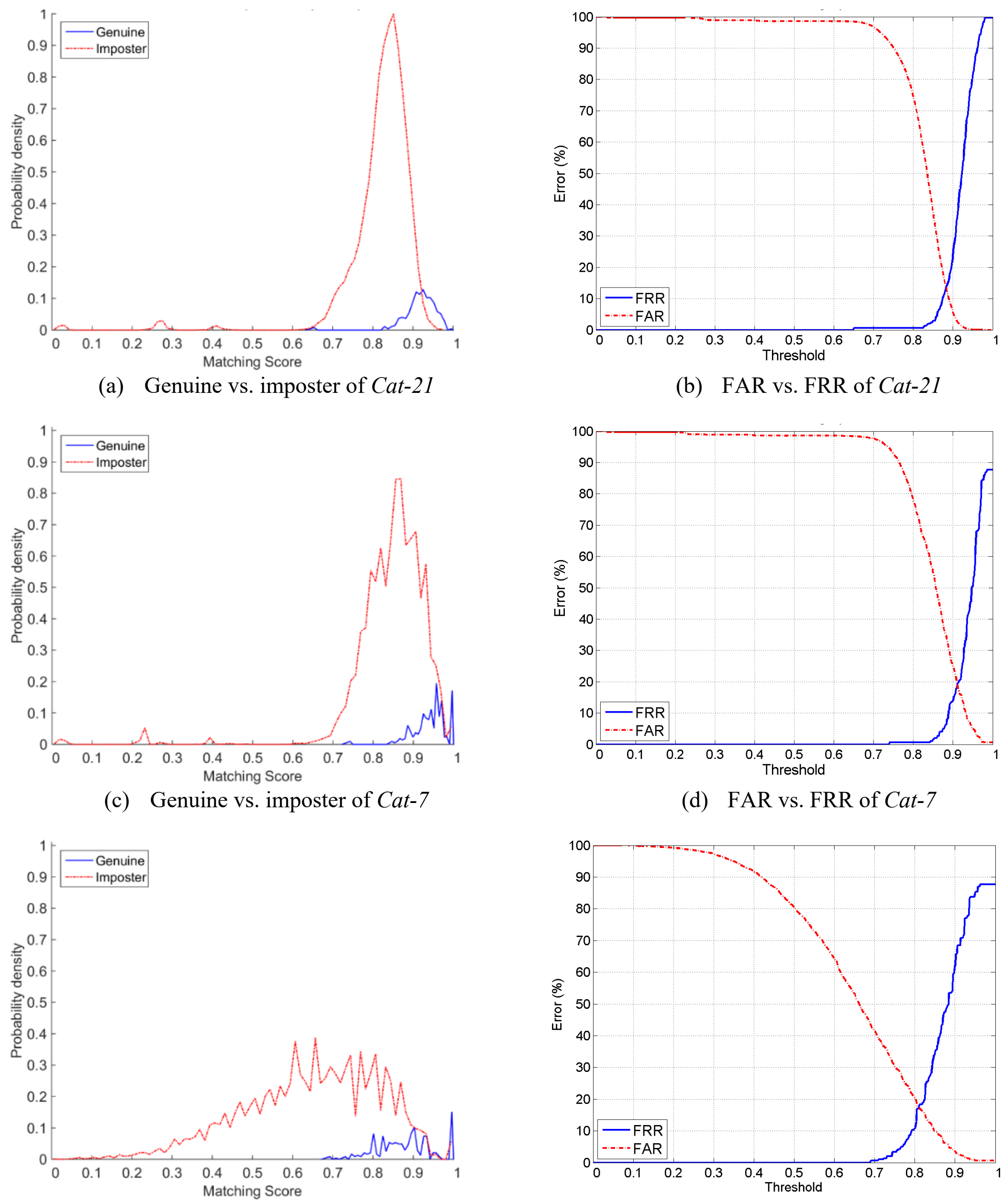


Figure 4.9: ROC performance of of soft clothing biometrics when used alone



(a) Genuine vs. imposter of *Cat-21* (b) FAR vs. FRR of *Cat-21*

(c) Genuine vs. imposter of *Cat-7* (d) FAR vs. FRR of *Cat-7*

(e) Genuine vs. imposter of *Cmp* (f) FAR vs. FRR of *Cmp*

Figure 4.10: Genuine vs imposter and error curves of clothing traits *alone* for identification

4.3 Conclusions

In this chapter a number of person identification experiments are conducted using categorical and comparative soft clothing biometrics in case when supplementing traditional or body soft biometrics and in case when they are used alone in isolation. The identification performance is reported and compared using different standard evaluation metrics.

Clothing characteristics can be utilized to convey effective descriptions and valuable combinations of soft biometrics. In some cases, human clothing descriptions might be the only observable attributes and a beneficial clue for identification.

The obtained performance results using semantic clothing attributes indicate a motivation for exploiting derived clothing traits, in fusion or even in isolation, to enrich human identification and re-identification. The proposed techniques enforce an identification task by using clothing descriptions, in supplement to traditional or body soft traits, and as the only biometrics. We enable the implicit use of the rich correlations of clothing attributes. We show that, as a proof-of-concept, we can recognize people using soft clothing traits and we need further to learn a soft attribute-centric for more powerful identification and retrieval in the future. The significant recognition improvement achieved by fusing clothing with soft biometrics can pave the way for further useful applications and motivates further research on different fusion approaches, investigating the capabilities of clothing attributes in supplementing another traditional biometrics.

Chapter 5

Retrieval by Soft Clothing Biometrics

- **Retrieval vs. Identification**

Since we have previously studied the identification capability of clothing in Chapter 4, this chapter aims further to investigate retrieval capabilities of soft clothing traits. As such, we validate and evaluate retrieval performance of clothing-based approaches through a number of standard evaluation metrics, afterwards we outline and discuss the experimental results. Biometric based retrieval can be described as a task that aims to identify an *unknown* subject by comparing and matching their biometric signature with those signatures enrolled in a database [27]. The distinction of retrieval against recognition that it concerns the ability to generalize to *unseen* data. For the sake of retrieval, the collected clothing annotations were divided into two sets: a *Query* set comprising one annotation per subject for each of the 128 subjects, which is used (as *unseen* data) to examine retrieval; and a *Training* set containing all the remaining annotations, which is used for training and feature selection processes. Fig. 5.1 clarifies the difference of biometric retrieval versus typical identification.

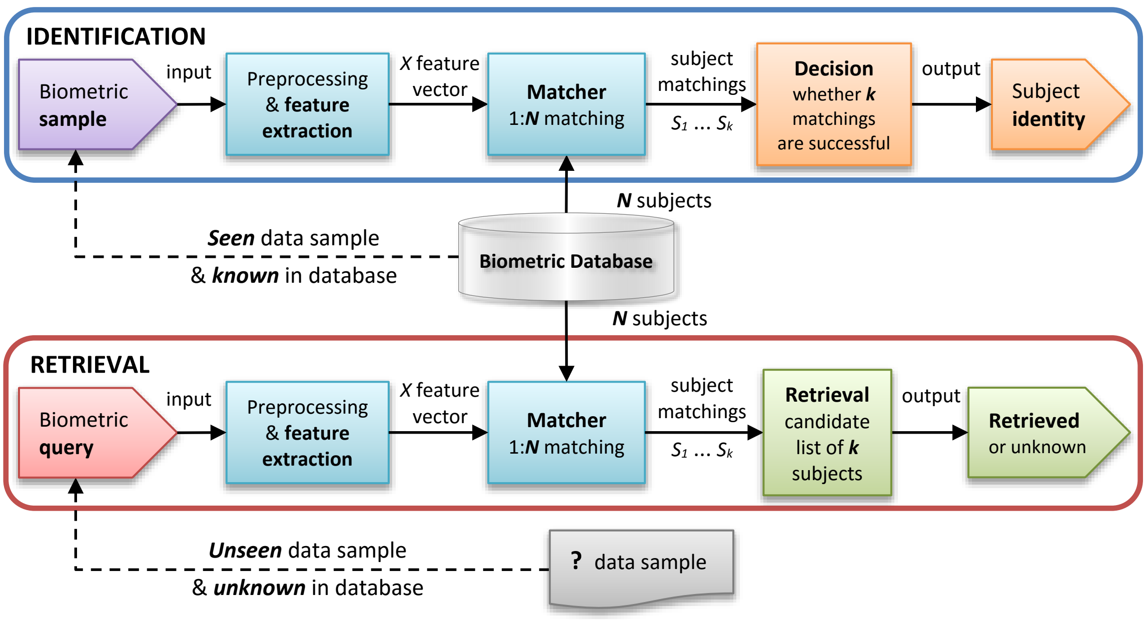


Figure 5.1: Barometric retrieval task versus identification task

5.1 Retrieval using Soft Clothing Traits

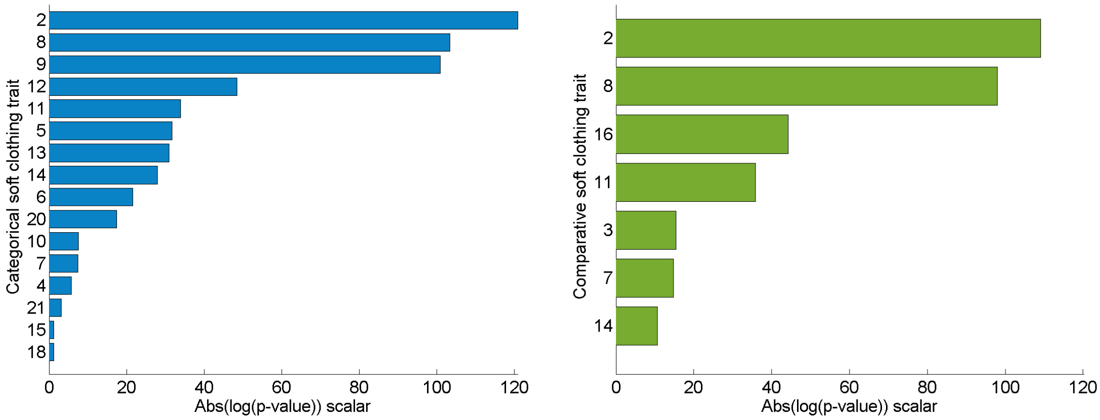
The main objective of this experimental work is to examine the proposed soft clothing approaches, described in Section 4, in retrieval and to explore their viability to supplement the performance of soft body biometrics. Our clothing analysis is used to enhance the retrieval of soft body descriptions (described in Section 2.2.1) in two different respects. Firstly, to enhance the performance of only the traditional soft traits (Age, Ethnicity, Sex, and Skin Colour) which were grouped as *Global* attributes as in [39]. Secondly, to enhance all the mentioned 23 soft body traits including the four traditional traits.

As feature subset selection, a new feature vector is formed by applying one-way ANOVA similar to the way described in Section 3.3. However, the F-ratio values of correlated and significant clothing traits are computed here using only soft clothing data annotations of the *Training* set to determine the most effective soft clothing traits for discrimination. After analysing traits separately, they were ranked as by p-value as shown in Fig. 5.2 and Table 5.1, then a minimum number of traits that achieve the best retrieval performance were selected. As such, the third feature vector is formed using a subset of the top five traits (attributes 2, 8, 9, 12, and 11), building a gallery called (*Cat-5*).

Table 5.1: Ordered list of soft clothing traits by their p-values

(a) Categorical trait		
Soft clothing biometrics	**F-ratio** ($df = 315$)	**P-value** ($p \leq 0.05$)
2. Head coverage	13.239	8.45E-53
8. Sleeve length	10.549	3.07E-45
9. Lower body clothing category	10.189	4.07E-44
12. Belt presence	4.600	2.46E-21
11. Leg length (of lower clothing)	3.478	5.07E-15
5. Upper body clothing category	3.324	4.25E-14
13. Shoes category	3.265	9.66E-14
14. Heel level	3.057	1.80E-12
6. Neckline shape	2.608	1.16E-09
20. Style category	2.326	6.96E-08
10. Shape (of lower clothing)	1.618	0.0013
7. Neckline size	1.608	0.0015
4. Hat	1.471	0.0081
21. Tattoos	1.214	0.1129
15. Attached object category	0.874	0.7910
18. Object in hand	0.874	0.7910

(b) Comparative trait		
Soft clothing biometrics	**F-ratio** ($df = 315$)	**P-value** ($p \leq 0.05$)
2. Head coverage	11.369	1.08E-47
8. Sleeve length	9.790	7.79E-43
16. Bag (size)	4.258	1.73E-19
11. Leg length (of lower clothing)	3.622	7.16E-16
3. Face covered	2.186	5.30E-07
7. Neckline size	2.140	1.03E-06
14. Heel level	1.852	6.07E-05



(a) Categorical traits (b) Comparative traits

Figure 5.2: Soft clothing attributes ordered according to p-values

All annotations in the *Query* set are excluded from the label data, and not reused in *Training* set and not involved in any training processes. The body *Training* set is used to derive for each subject a single feature vector consisting of 23 normalized average-labels. These average-labels are computed for a set of annotations provided by multiple users describing the 23 body traits of the same subject. A set of all computed feature vectors comprises a gallery called *softBody* to be tested separately in retrieval. Then each of the soft clothing galleries (*Cat-21*, *Cat-7, Cat-5,* and *Cmp*) is used to supplement *softBody* such that, each feature vector describing a single subject in *softBody* is concatenated to a corresponding feature vector describing the same subject in each of clothing galleries, resulting in: *softCat-21*, *softCat-7*, *softCat-5* and *softCmp* galleries.

Another gallery called *tradSoft* is derived from *softBody,* consisting of a four-trait feature vector per subject comprising only the four traditional soft descriptions (i.e. Age, Ethnicity, Sex, and Skin Colour). Likewise, *tradSoft* is extended to four versions supplemented by clothing to construct new galleries: *tradCat-21*, *tradCat-7*, *tradCat-5* and *tradCmp*. Query-vectors are normalized and reshaped according to the feature-vectors in a tested gallery to enable comparison and matching. Table 5.2 lists and describes the used soft biometric approaches (and produced galleries) of body traits and when combined with clothing traits.

The likelihood between every single query-vector and all subject-vectors in a gallery is estimated and retrieved, resulting in an ordered list of all subjects based on likelihood evaluated by the sum of Euclidean distance between query and gallery vectors. A number of standard performance evaluation methods are used to enable comparison between approaches from different perspectives. The same performance metrics used to evaluate identification in Section 4 are used here to evaluate retrieval comprising CMC, ROC, EER, AUC, and d'. In addition, with respect to all evaluation metrics, the overall performance is deduced for all approaches to rank them by overall score. Genuine and imposter distributions besides FAR and FRR errors are presented here as a further performance analysis. More information about theses metrics and performance evaluation have been provided in Section 4.1.1.

Table 5.2: Body and clothing approaches (galleries) used for testing and comparing retrieval performance

Clothing-based biometrics	**Description**
Cat-21	21 categorical soft clothing biometrics (binary/multiclass, and relative)
Cat-7	7 categorical soft clothing biometrics (relative)
Cat-5	The best 5 categorical soft clothing biometrics (feature subset selection by ANOVA)
Cmp	7 comparative soft clothing biometrics (relative)
Body-based biometrics	**Description**
tradSoft	4 categorical soft body biometrics (Age, Ethnicity, Sex, and Skin Colour)
softBody	23 categorical soft body biometrics including ***tradSoft***
Combined clothing and body biometrics	**Description**
tradCat-21	***Cat-21*** combined with ***tradSoft***
softCat-21	***Cat-21*** combined with ***softBody***
tradCat-7	***Cat-7*** combined with ***tradSoft***
softCat-7	***Cat-7*** combined with ***softBody***
tradCat-5	***Cat-5*** combined with ***tradSoft***
softCat-5	***Cat-5*** combined with ***softBody***
tradCmp	***Cmp*** combined with ***tradSoft***
softCmp	***Cmp*** combined with ***softBody***

5.1.1 Retrieval using Clothing and Traditional Soft Biometrics

The ROC performance of the examined approaches is compared in Fig. 5.4, where all clothing approaches but *tradCat-21* provide better retrieval accuracy and less error than *tradSoft*. Fig. 5.3 besides Table 5.3 report the CMC scores and the average-sum scores along different ranks, where Table 5.3 shows also the ROC analysis results for the traditional soft traits (Age, Ethnicity, Sex, and Skin Colour) and when adding soft clothing traits to them; all the best values are shown in bold. In all retrieval match scores, all clothing approaches enhance the retrieval performance of the traditional soft biometrics up to 9% on average at rank 128 achieved by *tradCat-5.* The best overall performance achieved by *tradCat-5* followed by *tradCmp* with slightly low performance and very close scores across all assessment metrics. However, *tradCmp* receives the best scores in d', the average score up to rank 10, and reaching the 100% accuracy at the smallest rank. Though *tradCat-7* attains the highest score at rank 1, it is then exceeded by *tradCat-5* and *tradCmp* along the rank increase. Fig. 5.5 shows genuine versus imposter distributions with their consequent FAR and FRR, representing the matching performance of clothing approaches.

Table 5.3. Performance metrics of traditional soft biometrics and when supplemented by clothing

Approach	Top rank	AVG sum match scores up to rank		100% accuracy achieved at rank	EER	AUC	*d'*	Performance overall rank
	=1	*=10*	*=128*					
tradSoft	0.14	0.27	0.838	106	0.196	0.146	1.611	4
tradCat-21	0.22	0.41	0.867	107	0.259	0.176	0.951	5
tradCat-7	**0.25**	0.48	0.900	95	0.208	0.132	1.419	3
tradCat-5	**0.25**	0.49	**0.925**	70	**0.167**	**0.105**	1.555	**1**
tradCmp	0.24	**0.50**	0.919	**67**	0.174	0.113	**1.768**	2

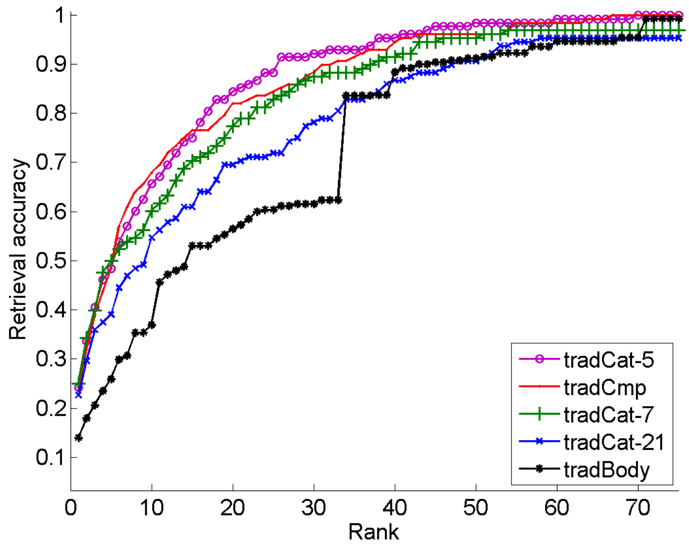


Figure 5.3: CMC (up to rank 75) of traditional soft biometrics and when supplemented by clothing

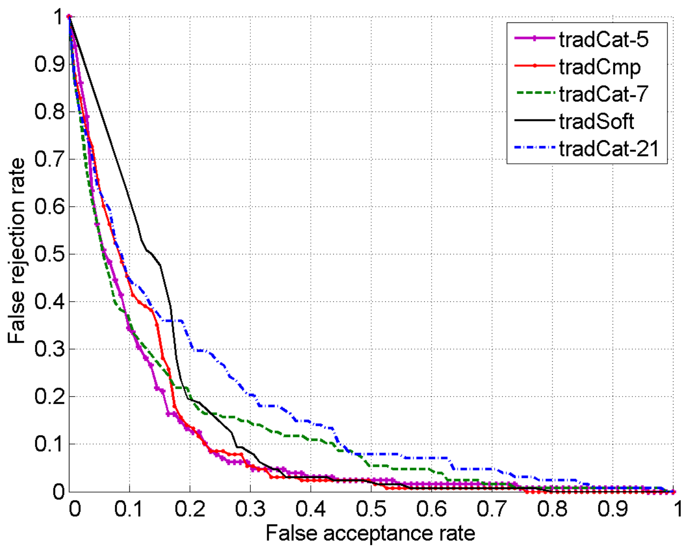
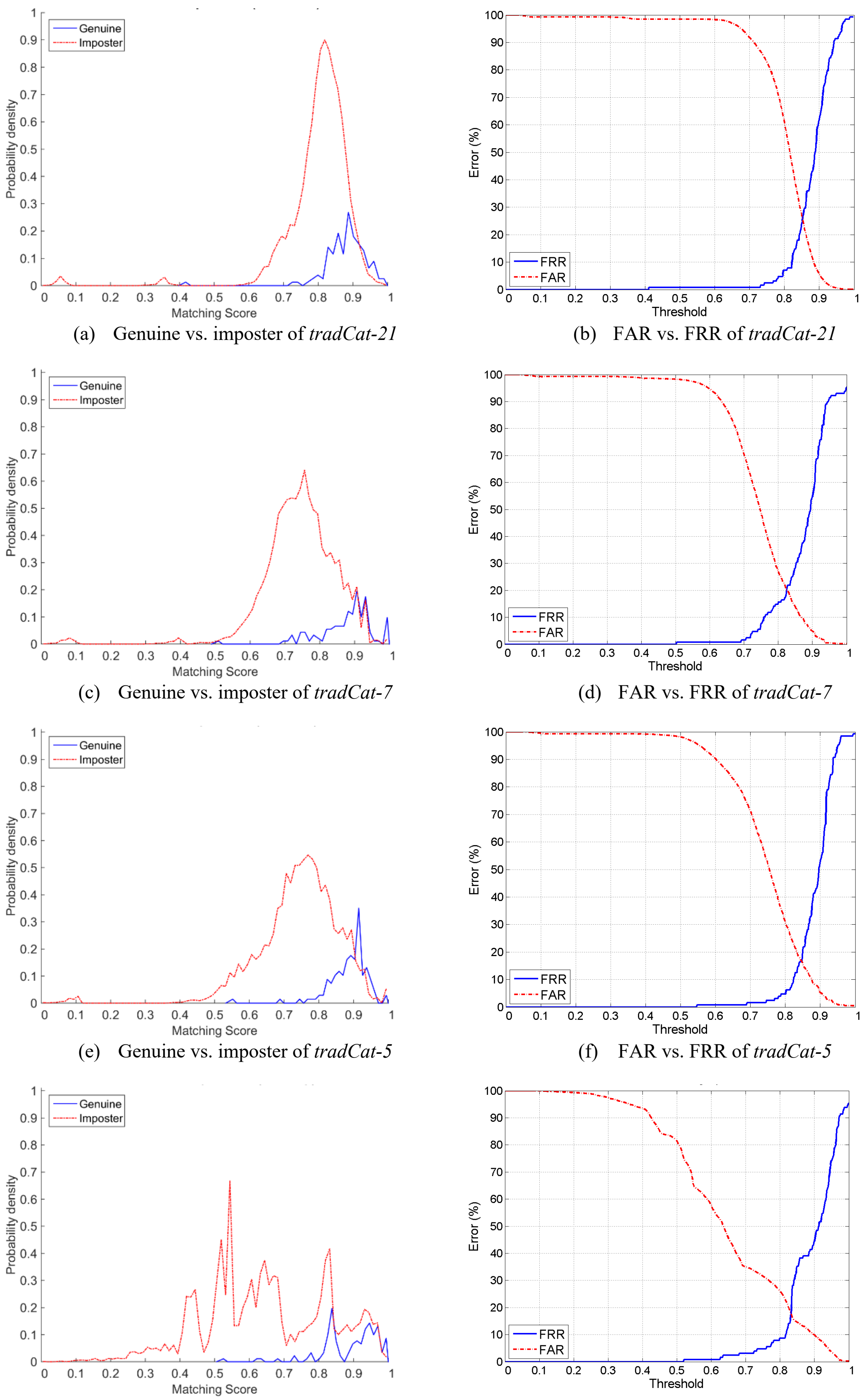


Figure 5.4. ROC performance of traditional soft biometrics and when supplemented by clothing



(a) Genuine vs. imposter of *tradCat-21*

(b) FAR vs. FRR of *tradCat-21*

(c) Genuine vs. imposter of *tradCat-7*

(d) FAR vs. FRR of *tradCat-7*

(e) Genuine vs. imposter of *tradCat-5*

(f) FAR vs. FRR of *tradCat-5*

(g) Genuine vs. imposter of *Cmp*

(h) FAR vs. FRR of *Cmp*

Figure 5.5: Genuine vs imposter and error curves of clothing traits added to *tradSoft* for subject retrieval

5.1.2 Retrieval using Clothing and soft body biometrics

Fig. 5.6 presents the CMC curves of the retrieval performance of soft body traits and compares the performance the clothing approaches. The figure shows the CMC curves up to rank 25 where the differences between compared approaches are more significant and can be appreciated. Table 5.4 provides all the produced metric results of CMC and ROC of soft body biometrics and the clothing approaches. The approaches *softCat-5*, *softCmp*, and *softCat-7* respectively gain a highest performance that improves retrieval performance of using soft body biometrics alone, *softCat-21* starts with a higher accuracy than *softBody* but provides a lower performance between rank 2 and 23, and then increase rapidly over all approaches.

The best overall performance is achieved by *softCat-5* as it yields the best scores in all evaluation measurements but one (i.e. the rank where 100% achieved) as can be observed in Table 5.4, which also receives the minimum error compared with other approaches as shown in Fig. 5.7 elaborating ROC performance. All clothing approaches started with better retrieval accuracy at rank 1 than the soft body traits, while *softCat-5* considerably increase the retrieval from only 67% to 82%. Although the inferiority of the *softCat-21* in all used evaluation criteria compared with its clothing-based counterparts, it is the first to reach the reach 100% at a minimum rank of 32. Fig. 5.8 shows for each clothing approach, genuine and imposter matching distributions and their matching errors, suggesting that *softCat-5* receives the least errors and has the lowest overlap between the genuine and imposter distributions, which is also compatible with the *d'* results showing *softCat-5* with the highest and best score.

Table 5.4: Performance metrics of soft body biometrics and when supplementd by clothing

Approach	**Top rank**	**AVG sum match scores up to rank**		**100% accuracy achieved at rank**	**EER**	**AUC**	***d'***	**Performance overall rank**
	=1	***=10***	***=128***					
softBody	0.668	0.900	0.988	56	0.094	0.039	2.379	4
softCat-21	0.703	0.884	0.987	**32**	0.107	0.049	1.836	5
softCat-7	0.742	0.922	0.990	40	0.112	0.038	2.294	3
softCat-5	**0.820**	**0.947**	**0.992**	47	**0.085**	**0.031**	**2.438**	**1**
softCmp	0.734	0.926	0.990	49	0.094	0.037	2.417	2

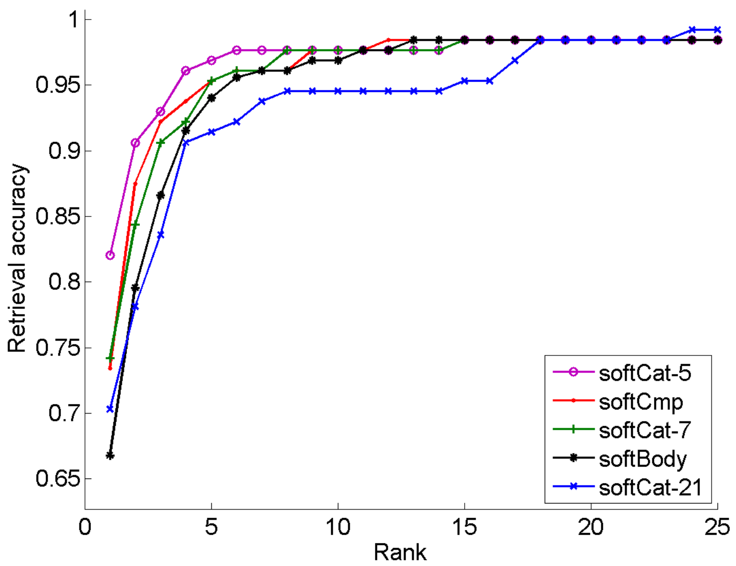


Figure 5.6: CMC performance (up to rank 25) of soft body biometrics and when supplementd by clothing

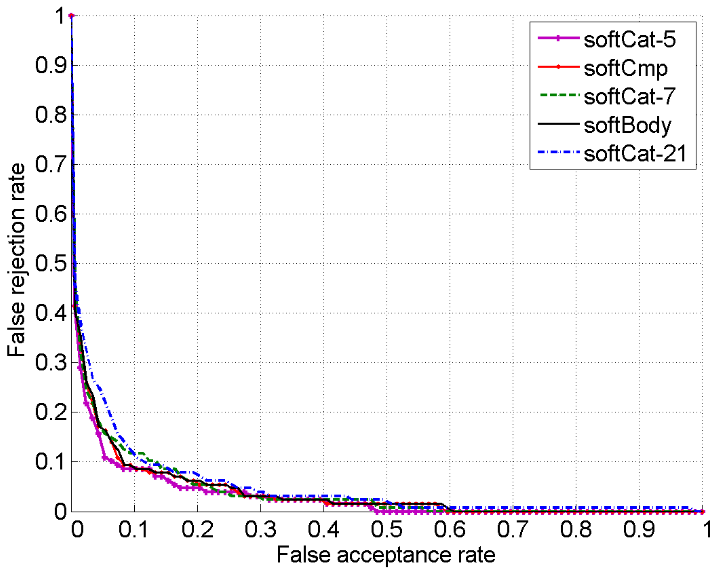
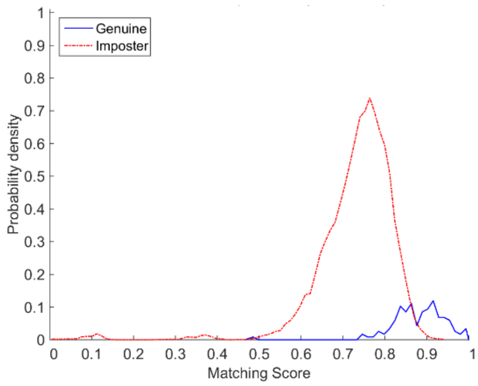
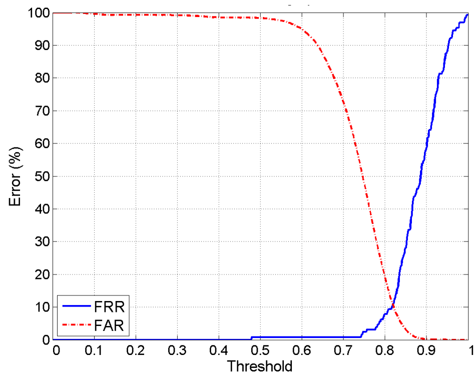


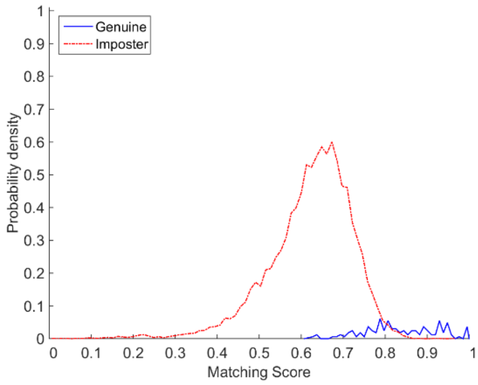
Figure 5.7: ROC performance of soft body biometrics and when supplemented by clothing



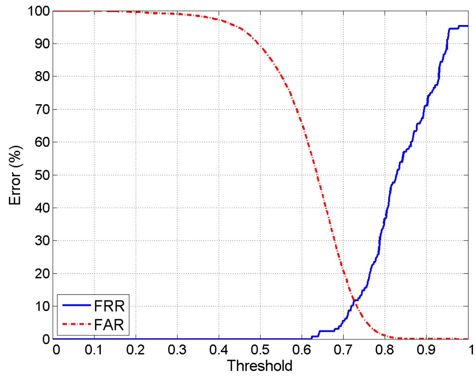
(a) Genuine vs. imposter of *softCat-21*



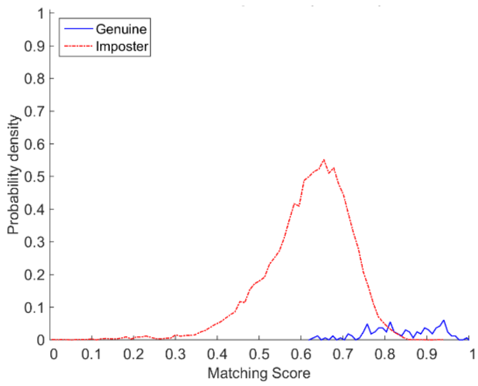
(b) FAR vs. FRR of *softCat-21*



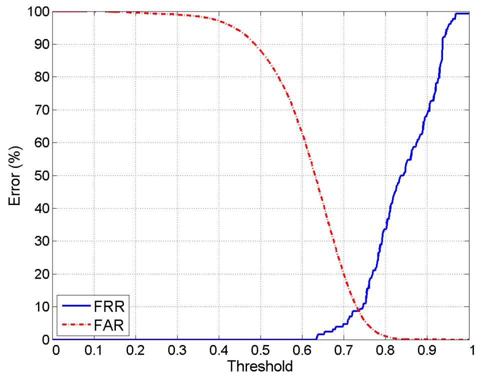
(c) Genuine vs. imposter of *softCat-7*



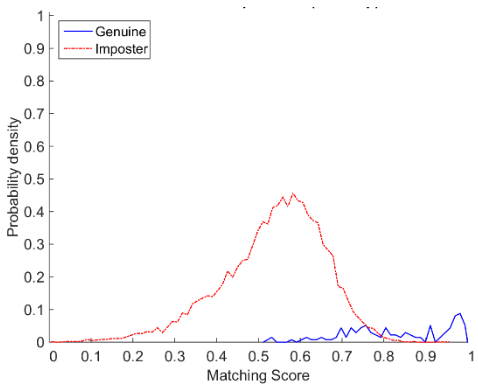
(d) FAR vs. FRR of *softCat-7*



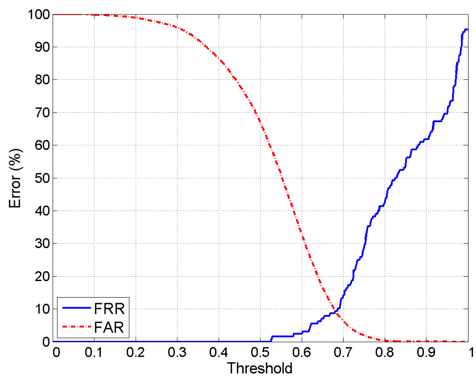
(e) Genuine vs. imposter of *softCat-5*



(f) *FAR vs. FRR of softCat-5*



(g) Genuine vs. imposter of *softCmp*



(h) FAR vs. FRR of *softCmp*

Figure 5.8: Genuine vs imposter and error curves of clothing traits added to *softBody* for subject retrieval

5.2 Viewpoint Invariant Retrieval using Soft Clothing Traits

5.2.1 Clothing Descriptions with Viewpoint Variation

The Soton Gait subset (described in Section 2.2) is used in this experiment, where the front-view images are used to collect clothing *Training* descriptions, and the side-view images are used to collect clothing *Query* descriptions. Fig. 5.9 shows a sample of both front and side view images of the same subject. It can be perceived that with the big change from front- to side-viewpoint, some clothing information are occluded or difficult to observe, such as belt presence and neckline size or shape.

Figure 5.9: *Front* and *side* samples used to obtain *Training* and *Query* annotations respectively

- **Clothing Data Annotation for *Training***

For training purposes, a set of categorical and comparative labels were collected from 27 users via a suitable web-form for each labelling type. All 128 front-view samples were labelled by multiple users, with one or more separate user annotations per subject describing the 21 categorical attributes. All subjects were compared using the seven relative (attributes 2, 3, 7, 8, 11, 14, and 16) by multiple users. Then to enrich the comparison data from the available number of collected comparisons, additional comparisons were inferred when two subjects were both compared with another same subject. A summary of collected *Training* data and the inferred comparisons data is shown in Table 5.5.

- ***Query* Data Annotation for Retrieval**

We designed a new web-form similar to the one (shown in Section 2.3) used for *Training* data collection. Fig. 5.10 demonstrates the new online form of the additional (Task 3), which shows subjects' side-view images instead of the front-view, and was used to collect a total of 107 *Query* annotations about six months later from 11 user annotators. Some of annotators were new to the system and had not participated in any annotation before, whereas the others had already participated in *Training* annotation. It was ensured that annotators were requested to annotate subjects different from those they had already observed before. This, in addition to the long time separation between *Training* and *Query* annotations, aimed to ensure that collected *Query* annotations were totally new descriptions and not affected by prior experience or annotator bias. The collected label data for side-view images were used only for performance analysis purposes. Table 5.5 summaries collected *Query* data.

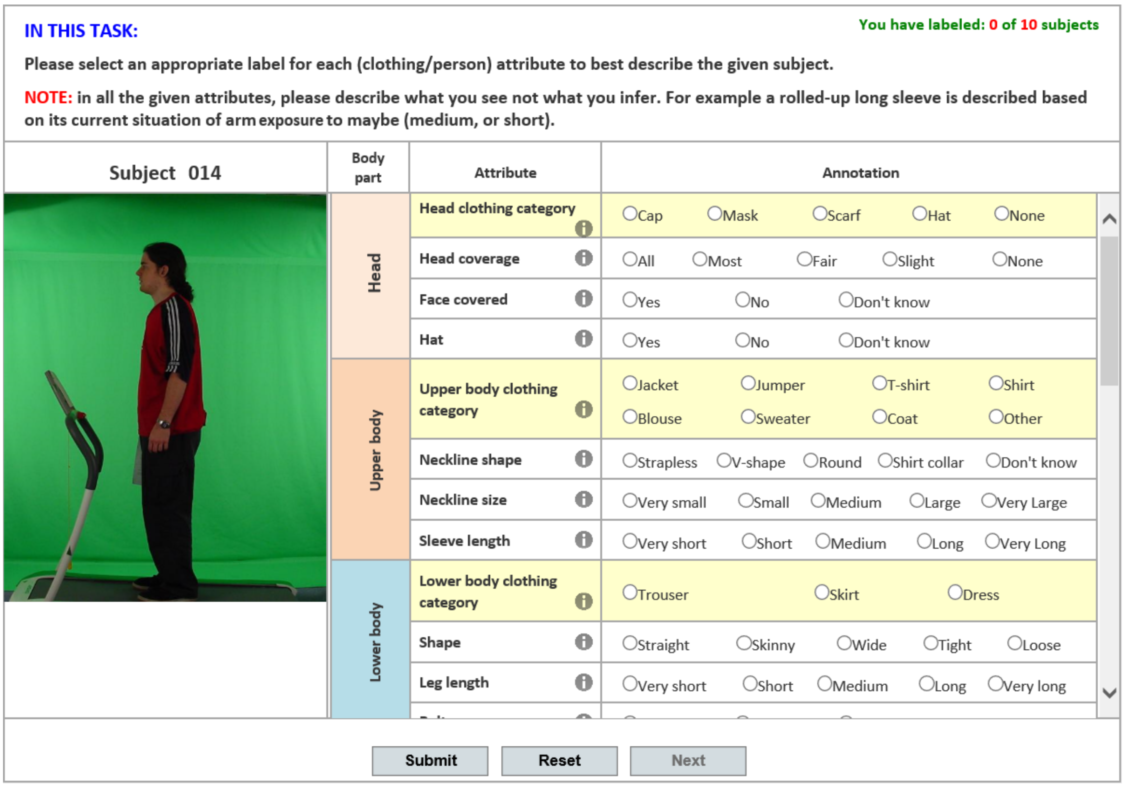


Figure 5.10: The online categorical annotation form of Task 3, describing side-viewpoint images

Table 5.5: The number of collected and inferred annotation data

***Training* data summary**	**Collected**	**Inferred**	**Total**
Total user annotations	316	N/A	316
Total user comparisons	317	556	873
Total attribute annotations	6636	N/A	6636
Total attribute comparisons	2219	3892	6111
***Query* data summary**	**Collected**	**Inferred**	**Total**
Total user annotations	107	N/A	107
Total attribute annotations	2247	N/A	2247

5.2.2 Invariant Soft Clothing Biometrics

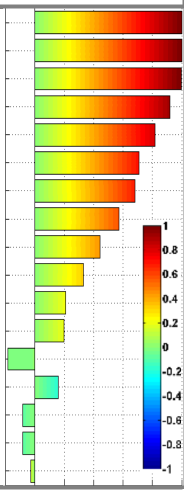
- **Categorical Clothing Traits (*Cat-N*)**

All collected categorical annotations in *Training* dataset are used to compose for each subject a single categorical feature vector of clothing descriptions. This feature vector comprises 21 soft clothing traits deduced from a set of labels provided by multiple users describing the same subject. Each soft clothing trait is computed as a normalized average-label for a single attribute. Hence, the resulting feature vectors for all subjects are gathered to construct a categorical-based gallery of soft biometric signatures, we refer to as *Cat-21*.

Table 5.6 presents an ordered list of Pearson's r value and p-value for correlated front-view and side-view categorical labels describing the same clothing attribute. A higher positive correlation value, closer or equal to 1, reflects a label's consistency and suggests that the attribute is invariant between different viewpoints. Our study on correlations between labels describing the same clothing attributes from the different viewpoints (front and side), can guide insight into which traits are more reliable and robust against viewpoint change. This in turn can lead to nominate a minimum number of effective categorical clothing traits to achieve enhanced invariant subject retrieval. As such, we reshape from *Cat-21* a second categorical gallery called *Cat-6*, comprising for each subject a feature vector composed of only a subset of the top six correlated traits (2, 4, 1, 9, 8, and 11) , derived via ANOVA, shown in Table 5.6.

Table 5.6: Correlated *front* and *side* view labels per same-attribute

Soft clothing trait	P-value ($p \leq 0.05$)	Pearson's r
2. Head coverage	0.000	1.000
4. Hat	0.000	1.000
1. Head clothing category	0.000	1.000
9. Lower body clothing category	5.13E-45	0.922
8. Sleeve length	4.71E-27	0.819
11. Leg length (of lower clothing)	1.37E-17	0.709
5. Upper body clothing category	4.97E-16	0.684
13. Shoes category	1.13E-10	0.573
12. Belt presence	1.74E-06	0.443
7. Neckline size	0.001	0.331
20. Style category	0.030	0.210
6. Neckline shape	0.041	0.198
21. Tattoos	0.057	-0.184
10. Shape (of lower clothing)	0.107	0.157
15. Attached object category	0.396	-0.083
18. Object in hand	0.396	-0.083
14. Heel level	0.778	-0.028



- **Comparative Clothing Traits (*Cmp*)**

All comparisons in the *Training* dataset are used to derive for each subject a single feature vector of comparative clothing descriptions. Each value of the feature vector is supposed to be a relative measurement reflecting the degree of presence of a single comparable attribute. Thus, all collected comparative annotations need to be anchored, per attribute, to define these invariant relative measurements for each subject, and to arrange a list of ordered subjects with respect to a single attribute. To derive desired relative measurements and to achieve ordering for all subjects per attribute, we use the same Ranking SVM method described in Section 2.4.

Hence, from the *Training Cat-21* gallery, we form for each subject a feature vector using a subset of seven categorical traits, which are the only seven relative clothing attributes, shown in bold in Table 2.1 (attributes 2, 3, 7, 8, 11, 14, and 16). The resulting 7-value feature vectors of all subjects are used to learn seven optimal ranking functions for the seven relative attributes. The weighting of each function is derived using the formulation in Eqn. (2.2). The desirable per attribute ordering of all subjects is deduced from w. Then by Eqn. (2.1) using only the *Training* annotation data, each value of w is used to map each of 7-value feature vectors to a corresponding vector of seven relative measurements (i.e. comparative traits) describing a single subject. Eventually, a set of all mapped feature vectors is used to build a gallery of comparative clothing descriptions for all subjects, referred to as *Cmp.*

5.2.3 Viewpoint Invariant Retrieval Methodology

In the context of this research, biometric based retrieval is achieved such that, a query description of an unknown subject is used to probe a gallery by comparing and matching their biometric signature with all those signatures enrolled in the gallery. Subject retrieval can be distinguished from identification by which it is a task aims to identify an *unknown* subject using their biometric signature to find a match in those signatures enrolled in a database [27], in other words, it concerns the ability to generalize to unseen data. Therefore, retrieval tends to be more challenging and beneficial for extensive biometric applications. It is worth emphasising that a subject *Query* description is obtained from an annotator different from those annotators who provided the *Training* descriptions for the same subject. Moreover, a subject *Query* description is provided to describe an image different from the image of the same subject used for *Training* descriptions, and with different viewpoint (i.e. side view), where a number of clothing attributes are likely to be occluded or difficult to observe. It appears that such a context of subject retrieval provides extra challenges emulating some real-life cases in their conditions and complexity.

Soft body descriptions are available for Soton database, which were previously collected [39] such that, each of 115 individuals was labelled by multiple users' describing 23 soft body traits. Here, the performance of subject retrieval using these soft body biometrics alone is used as a baseline to be compared with their performance when supplemented by soft clothing biometrics. Soft clothing biometrics are examined in enhancing subject retrieval of two groups of soft body traits. The first group consist of only four (global) traditional soft traits (Age, Ethnicity, Sex, and Skin Colour), which is used to construct the gallery (*tradSoft*) comprising all feature vectors of 4 normalized average-labels. The second group comprises 17 soft body traits, observable from front and side views, including the former four traditional traits, which is used to construct the gallery (*softBody*) comprising all feature vectors of 17 normalized average-labels. Table 5.7 provides a list of traits under each soft body or clothing gallery, highlighting selected subset of traits used throughout the invariant retrieval experiments.

Table 5.7: A list for each body or clothing gallery, highlighting selected traits used for invariant retrieval

ID	*softBody*	*tradSoft*	*Cat-21*	*Cat-6*	*Cmp*
1	Age	Age	Head clothing category	Head clothing category	
2	Arm Length		Head coverage	Head coverage	Head coverage
3	Arm Thickness		Face covered	Face covered	Face covered
4	Chest		Hat	Hat	
5	Ethnicity	Ethnicity	Upper body clothing category	Upper body clothing category	
6	Facial Hair Colour		Neckline shape	Neckline shape	
7	Facial Hair Length		Neckline size	Neckline size	Neckline size
8	Figure		Sleeve length	Sleeve length	Sleeve length
9	Hair Colour		Lower body clothing category	Lower body clothing category	
10	Hair Length		Shape (of lower clothing)	Shape (of lower clothing)	
11	Height		Leg length (of lower clothing)	Leg length (of lower clothing)	Leg length
12	Hips		Belt presence	Belt presence	
13	Leg Length		Shoes category	Shoes category	
14	Leg Direction		Heel level	Heel level	Heel level
15	Leg Thickness		Attached object category	Attached object category	
16	Muscle Build		Bag	Bag	Bag (size)
17	Neck Length		Gun	Gun	
18	Neck Thickness		Object in hand	Object in hand	
19	Proportions		Gloves	Gloves	
20	Sex	Sex	Style category	Style category	
21	Shoulder Shape		Tattoos	Tattoos	
22	Skin Colour	Skin Colour			
23	Weight				

Categorical and comparative soft clothing galleries (*Cat-21*, *Cat-6*, and *Cmp*) are used along with the two soft body galleries (*tradSoft*, and *softBody*) to compose six combined galleries enabling to examine and evaluate the capability and performance of soft clothing biometrics in subject retrieval. Table 5.8 lists and describes the produced soft biometric galleries of body traits and when combined with clothing traits. The combination between 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 set of 107 annotations are excluded from the obtained soft body data to be used as query-vectors. For each combined gallery, corresponding query-vectors are normalized then reshaped and concatenated according to the feature vectors in a tested gallery to enable matching and retrieval. Based on the *k* nearest neighbour 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); and the Decidability Index (d'). With respect to all metrics, all approaches are ranked by overall performance. Detailed performance evaluation methodology have been already presented in Section 4.1.1.

Table 5.8: Body and clothing approaches (galleries) for retrieval experiments

Body-based soft biometrics	
tradSoft	4 categorical soft body biometrics (Age, Ethnicity, Sex, and Skin Colour)
softBody	17 categorical soft body biometrics including ***tradSoft*** (observable from front/side views)
Combined clothing & soft body biometrics	
tradCat-21	21 categorical clothing traits combined with ***tradSoft***
softCat-21	21 categorical clothing traits combined with ***softBody***
tradCat-6	The best 6 categorical clothing traits with ***tradSoft***
softCat-6	The best 6 categorical clothing traits with ***softBody***
tradCmp	7 comparative clothing traits combined with ***tradSoft***
softCmp	7 comparative clothing traits combined with ***softBody***

5.2.4 Invariant Retrieval by Clothing with Traditional Soft Biometrics

Here, clothing descriptions are used to supplement the traditional soft traits (Age, Ethnicity, Sex, and Skin Colour) in subject retrieval, the experimental results detailed in Table 5.9 and the CMC curves shown in Fig. 5.12 suggest that the retrieval performance is considerably and consistently enhanced by clothing traits in *tradCmp* and *tradCat-6* of clothing approaches. The retrieval accuracy of *tradSoft* at rank 1 is effectively enhanced by all clothing approaches up to 17% in the best case yielded by *tradCat-6*, and the average score up to rank 10 of *tradSoft* is also improved with a significant increase ranging from 9% to 28%. *tradCmp* is the best in terms of the decidability index and achieves the highest overall rank, indicating a better separation between genuine and imposter distributions and more successful discrimination between subjects. *tradCat-6* is the first to achieve 100% at rank 58. Fig. 5.13 illustrates that *tradCmp* followed by *tradCat-6* receive the best ROC scores with less errors outperforming *tradSoft* and *tradCat-21*. The performance of *tradCat-21* is low compared with the other clothing approaches, though it attains some recognition capability improving up to rank 34. Fig. 5.14 shows genuine versus imposter distributions with their consequent FAR and FRR, representing the matching performance of clothing approaches.

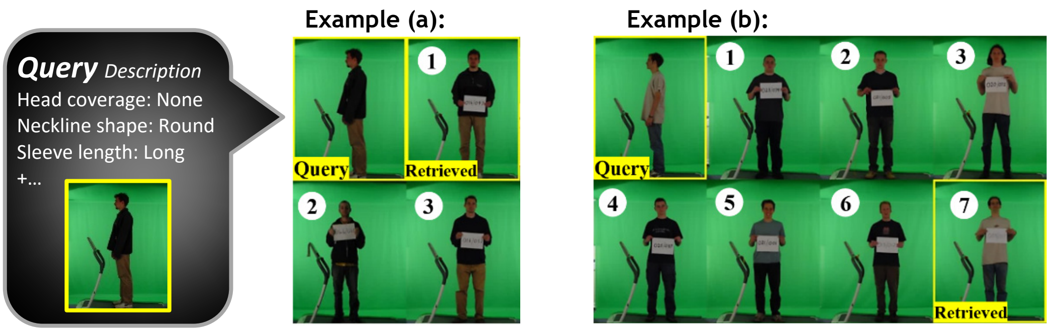


Figure 5.11: Retrieval query example using *tradCmp*, Example (a): a subject correctly retrieved at rank 1, Example (b): a subject retrieved but at rank 7

Fig. 5.11 shows two query examples of subject retrieval using *tradCmp* approach achieving the highest overall retrieval performance. In both examples, the top left corner image represents a side-view image used to derive the side-query descriptions. The remaining numbered front-view images represent the top k retrieved subjects from the test gallery, ordered based on their similarity to the query-description, where the query image and the correctly retrieved subject are bordered in yellow. In the first example on the left a query subject was correctly retrieved at rank 1, whereas in the second example on the right another subject was retrieved only at rank 7. In the second (right) example, it can be observed that all retrieved subjects are very similar in their clothing such as sleeve and leg length, and these similarities are correctly reflected by the match. It appears that in such a case, the strong similarities in comparative clothing traits and in the four traditional biometrics, may result in a confusion between such similar subjects. Despite of the desirable objective of retrieving the correct subject as the top match (rank 1), the retrieval of the correct subject within a small list (e.g. 10 subjects) appears reasonably successful and certainly will be useful to narrow the search. So, retrieval may not always answer the question "is the top match correct?" but instead could answer "is the correct answer in the top *k* matches?"

Table 5.9: Performance metrics of traditional soft biometrics and when add clothing

Approach	Top rank	AVG sum match scores up to rank		100% accuracy achieved at rank	EER	AUC	*d'*	Overall rank
	=1	*=10*	*=128*					
tradSoft	0.176	0.347	0.872	73	0.183	0.127	1.882	3
tradCat-21	0.234	0.443	0.865	104	0.302	0.198	0.882	4
tradCat-6	**0.346**	**0.617**	0.944	**58**	0.144	0.090	1.223	2
tradCmp	0.308	0.607	**0.946**	69	**0.113**	**0.077**	**2.006**	**1**

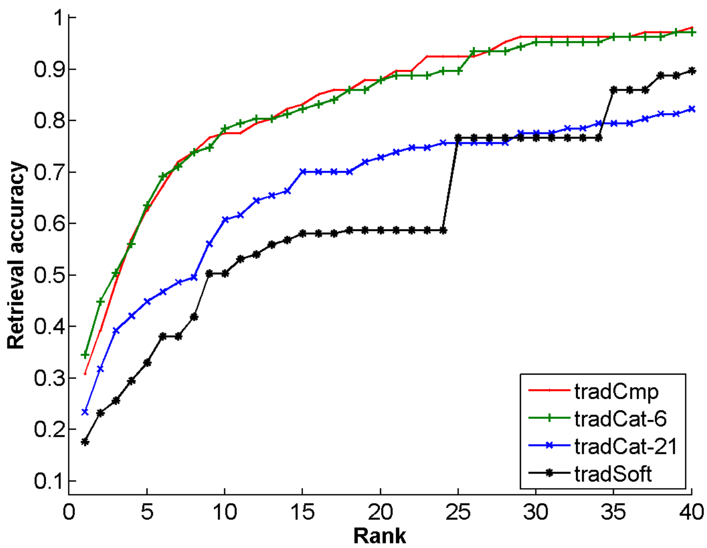


Figure 5.12: CMC of traditional soft biometrics and when add clothing

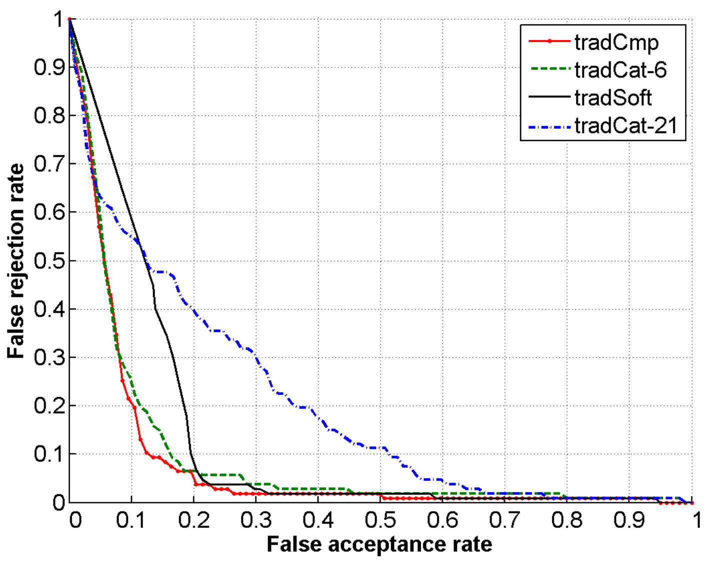
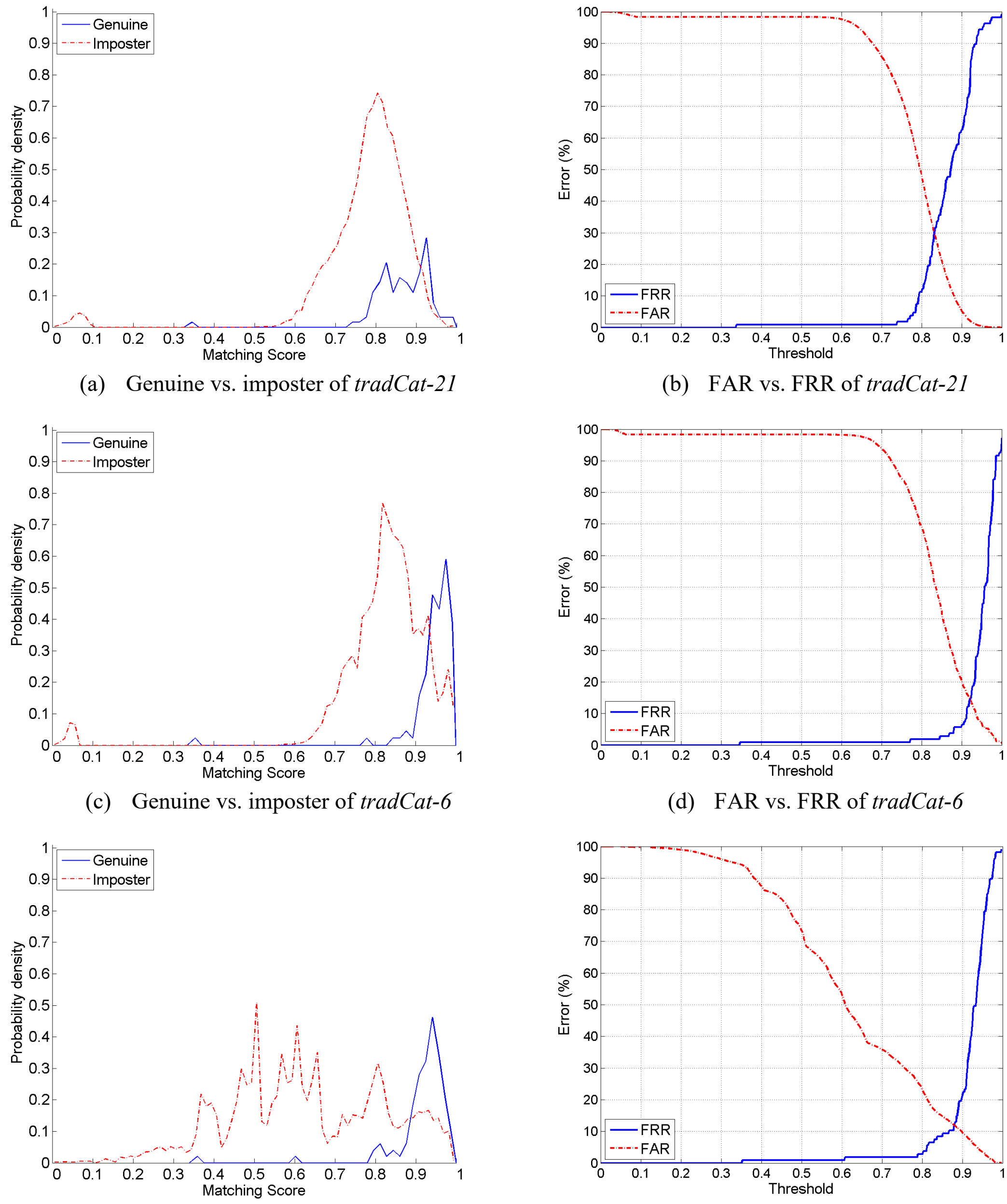


Figure 5.13: ROC of traditional soft biometrics and when add clothing



(a) Genuine vs. imposter of *tradCat-21*
(b) FAR vs. FRR of *tradCat-21*
(c) Genuine vs. imposter of *tradCat-6*
(d) FAR vs. FRR of *tradCat-6*
(e) Genuine vs. imposter of *tradCmp*
(f) FAR vs. FRR of *tradCmp*

Figure 5.14: Genuine vs imposter and error curves of clothing traits with *tradSoft* for invariant retrieval

5.2.5 Invariant Retrieval by Clothing with Soft body Biometrics

Clothing descriptions are reused here to supplement the front/side observable 17 soft body descriptions including the four traditional traits (see Table 5.8). Table 5.10 reports metric scores of all approaches, and Fig. 5.15 compares the CMC performance (up to rank 40) of these approaches. The clothing approaches *softCat-6* and *softCmp* respectively provide the highest performance, while *softCat-6* gains the best scores in all evaluation metrics but d' and EER, and outperforms all approaches as also reflected by the ROC in Fig. 5.16.

Fig. 5.17 shows for each clothing approach, genuine and imposter matching distributions and their matching errors. Further performance analysis of *softCat-6* is shown in Fig. 5.17-c which demonstrates estimated distributions of match scores in terms of intra-class (genuine) and inter-class (imposter), and Fig. 5.17-d represents the EER and trade-off between two errors FAR and FRR along a set of thresholds varying from 0 to 1. The rank 1 retrieval of *softBody* is obviously enhanced when adding clothing from about 86% to 92% by *softCmp*, and to 94% by *softCat-6*, given in Table 5.10. As such, the clothing analysis can effectively augment soft body descriptions. This enhanced performance is reflected in the class distributions and errors shown in Fig. 5.17 which confirm the potency *softCat-6* and *softCmp* labels. Since *Query* descriptions have been acquired using side-view images, some items are difficult to observe or occluded such as neckline shape or size and belt presence. As *softCat-21* consists of all clothing traits including those affected traits, this can produce, for affected traits, inconsistent descriptions and undesirable increase in within-class variance. That appears to be a reason for the low performance compared with the other approaches. Another reason could be the noise caused by adding a large number of clothing traits to another large number of body traits.

Table 5.10: Performance metrics of soft body biometrics and when add clothing

Approach	Top rank	AVG sum match scores up to rank		100% accuracy achieved at rank	EER	AUC	*d'*	Overall rank
	=1	*=10*	*=128*					
softBody	0.868	0.962	0.996	30	**0.064**	0.015	**3.442**	3
softCat-21	0.729	0.899	0.987	39	0.108	0.042	1.802	4
softCat-6	**0.943**	**0.985**	**0.998**	**27**	0.070	**0.012**	3.117	**1**
softCmp	0.916	0.981	0.997	30	0.070	0.014	3.406	2

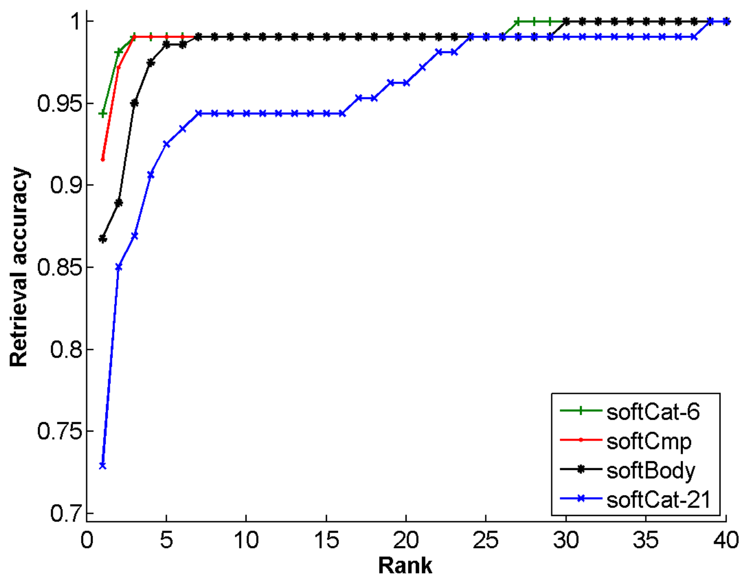


Figure 5.15: CMC (rank 40) of soft body biometrics and when add clothing

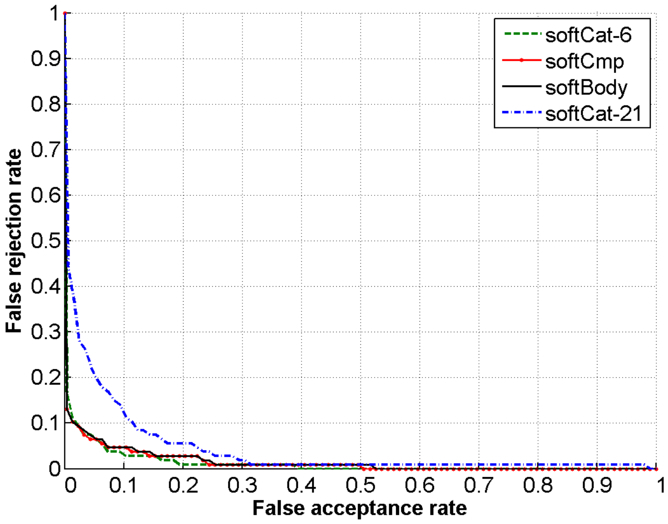
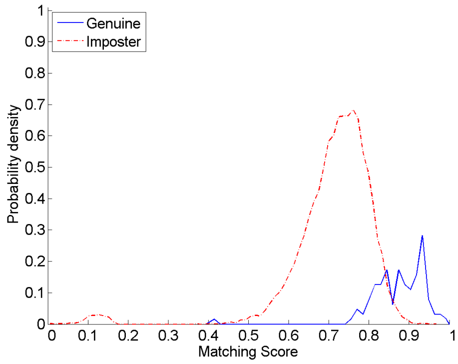
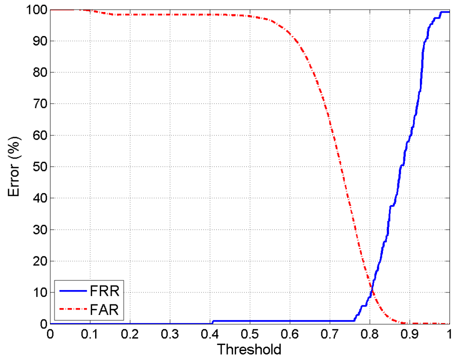


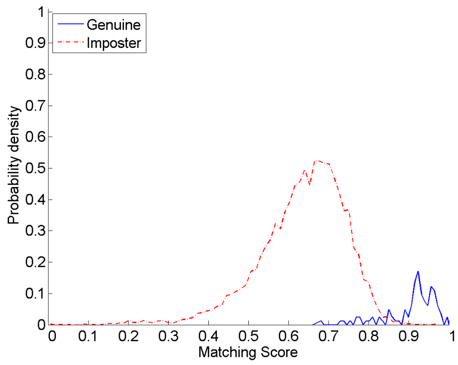
Figure 5.16: ROC of soft body biometrics and when add clothing



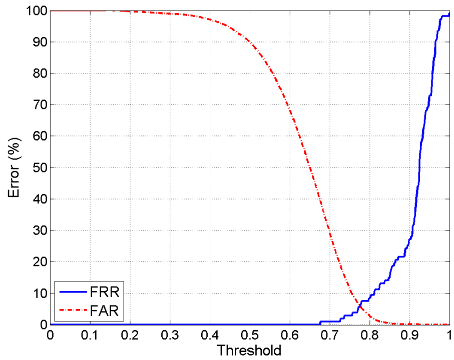
(a) Genuine vs. imposter of *softCat-21*



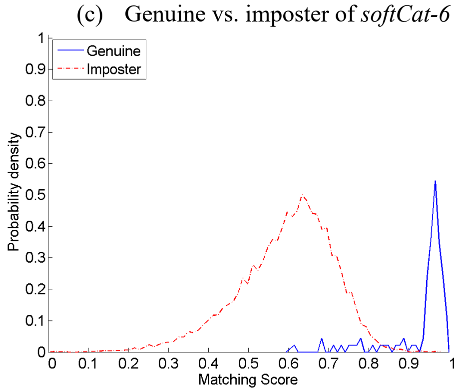
(b) FAR vs. FRR of *softCat-21*



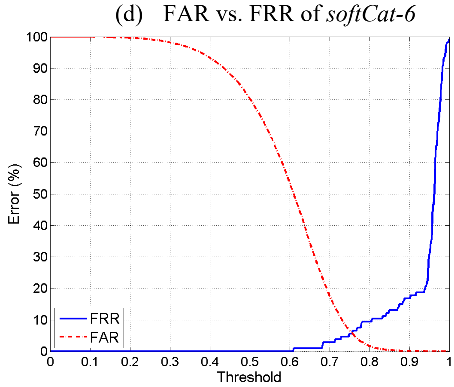
(c) Genuine vs. imposter of *softCat-6*



(d) FAR vs. FRR of *softCat-6*



(e) Genuine vs. imposter of *softCmp*



(f) FAR vs. FRR of *softCmp*

Figure 5.17: Genuine vs imposter and error curves of clothing traits with *softBody* for invariant retrieval

5.3 Conclusions

In this chapter biometric retrieval has shown how soft clothing biometrics can be generalized for unseen data and effectively used for enhanced subject retrieval. The experimental results of retrieval by soft clothing traits highlight a potentially valuable addition to the field of soft biometrics. This can lead to new and useful enhanced biometric applications and systems, using soft clothing biometrics for various purposes including subject search, retrieval, identification, and re-identification. Our evaluations indicate that such clothing characteristics can be associated in soft biometric signatures, achieving successful subject retrieval.

Soft clothing traits are further tested and evaluated in more challenging biometric scenarios, exploring their capabilities in achieving viewpoint invariant subject retrieval. Thus more realistic retrieval experiments are conducted aiming to increase the capability of recognition and retrieval from surveillance imagery suffering from several high variability in many conditions like viewpoint variation. The results show that soft clothing biometrics can be effectively used to supplement soft body biometrics and they are capable of significantly increasing retrieval performance.

Using comparative clothing traits and a small set of categorical clothing traits, which are highly correlated between multiple viewpoints, yields more accurate human descriptions and is more likely to achieve better performance in viewpoint invariant subject retrieval. The performance when adding clothing surpasses traditional and body soft traits when used in isolation. Although image-view variation can affect the observation of some clothing traits, the perception of other clothing traits remains invariant to viewpoint change and retains the capability for subject retrieval. Here, the new clothing descriptions are shown to be relatively immune change from front to side view.

Chapter 6

Vision-based Automatic Soft Clothing Attribute Extraction

6.1 Automatic Soft Clothing Attributes

In the majority of existing vision-based research, soft clothing attributes are suggested by either mining web fashion data [14, 22] or manual designs allowing to suggest more visually separable soft attributes to be more suited for biometric use [9]. Unlike most existing vision-based approaches concerning clothing attributes for fashion search/retrieval or 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* and *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 6.1 only ten soft attributes, shown in bold, are relative and can be described by a categorical form of labelling (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 labelling (such that *Lower Brightness*: 'Much darker, 'Darker', 'Same', 'Bright', 'Much brighter').

For non-relative attributes, each categorical label is coded by an integer value representing the textual expression of the label. For relative attributes, to better reflect the degree-of-strength and the degree-of-comparison via categorical and comparative labels, we define bipolar scales in a way inspired from an early analysis that proposed to characterize human traits for whole-body descriptions [82]. Each of the ten relative attributes is formed as a 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 colours 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 6.1 shows the proposed set of semantic attributes describing clothing with regards to three different aspects: *Overall body*, *Upper body*, and *Lower body*, where 'A' letter denotes the word *Automatic* and used along with a number to represent an automatic attribute ID here and so forth. From computer-vision perspective, these 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 colours, brightness, and contrast; and the third group of (A10 and A16), characterizing clothing patterns.

Table 6.1: Automatic soft clothing attributes and corresponding labels

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

6.2 Automatic 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. 6.2 for both front- and side-view images; more examples with detailed processes are given in the appendix. The process of this approach comprises four main phases: preprocessing, Analysing appearance, Analysing colours, and Analysing patterns. Fig. 6.1 illustrates an overview of our experimental framework for automatic soft clothing attributes extraction and labelling.

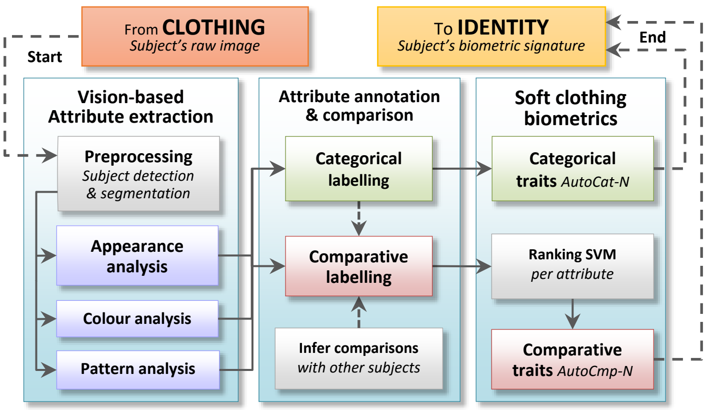


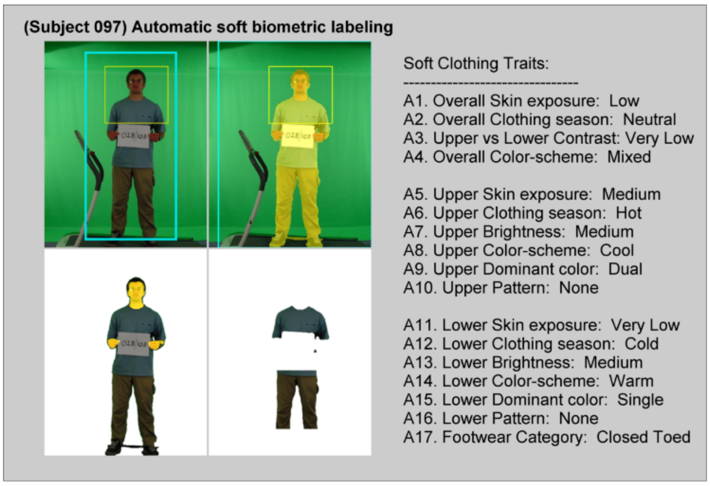
Figure 6.1: Overview of experimental framework of automatic clothing attributes extraction and labelling

6.2.1 Preprocessing

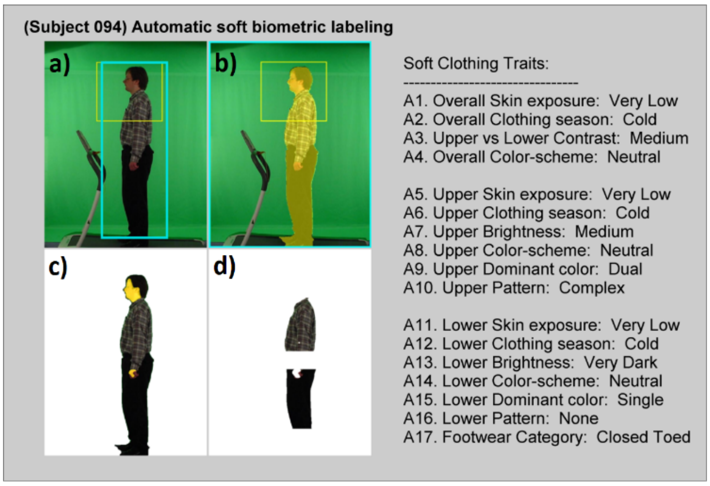
The proposed approach is designed to enforce soft clothing attribute extraction on subject images with different viewpoints (i.e. front and side). Therefore, a combination of body and face detection is considered for better human detection results. This enables a more accurate and reliable detection of the subject in an image.

Holistic human body detection is achieved using a deformable model-based detector [106], and a cascade classifier based on Haar-like features is used for face detection [107].

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. 6.2). Hence, the body box may be realigned to the face box and vice versa.



(i) Automatic clothing annotation for a front-view subject image



(ii) Automatic clothing annotation for a side-view subject image

Figure 6.2: A summary screen of a set of all 17 categorical soft clothing attributes automatically extracted and labelled to describe (i) front- and (ii) side-viewpoint subject images. Sub-images (a), (b), (c), and (d) show some applied steps of automatic vision-based image processing, including detection, analysis, segmentation, etc.

Afterwards, the final estimated region containing of the face/head (if no detectable face) is used to initialize a starting region of a foreground highlighting method [108], which employs the Grabcut technique for foreground object extraction [109] by learning foreground and background colour 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 colour models, shown in Fig. 6.3, to attain accurate full body segmentation. These colour models are generally designed with similar concepts as in [108] but with the following three characteristics:

First, the colour 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 regions.

Second, the colour 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 but maybe containing some foreground, and U being a neutral region equally free to form either foreground or background.

Third, the initial foreground colour model F is rescaled to mostly fit within the person's body captured from the side viewpoint as well as the front.

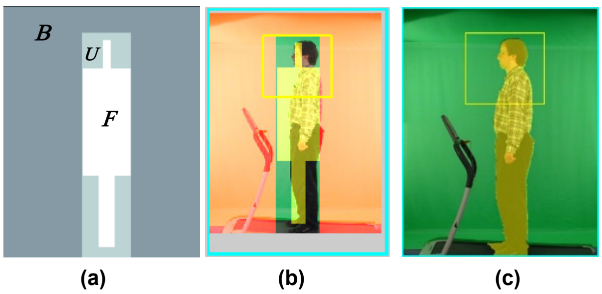


Figure 6.3: (a) Colour models used to initialize the Grabcut person extractor, (b) Resized and fitted colour models initializing a foreground highlighting method in a tested subject image. (c) Resulting highlighting for subject segmentation

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 6.1, 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 [110].

6.2.2 Analysing Appearance

Skin exposure soft attributes (A1, A5, and A11) and clothing season 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 [111] is used to compute a pixel-level skin likelihood probability. Then we apply three skin tone detectors, in three colour spaces, HSV, YCbCr, and RGB, by defining the ranges of colour values for each channel, which are likely to represent the variety of skin colours. The model-based skin detector is sensitive to shaded regions and negligible details in body-level detection (like eyes, eyebrows, 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-colour-based detectors. The YCbCr detector was found to be the most effective and coherent in detecting most skin colour 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-labelling, 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 colours and patterns analysis.

6.2.3 Analysing Colours

Clothing colours are analysed 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 colours of upper and lower parts (A9, and A15); and colour-scheme category of overall, upper, and lower clothing (A4, A8, and A14). The RGB upper/lower clothing segment is converted to 256 indexed colour space and then approximated to 64 indexed colours using a minimum variance quantization, to be converted further to grayscale colour space for brightness analysis. This process is useful for minimizing clothing wrinkles and shadow effects. We design and use a five-scale colour-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 colours are detected and classified, each into (one of the five) brightness groups (or levels) as applicable. Then these K colour groups are used along with a voting formulation V defined in Eqn. (6.1), to nominate the most appropriate label L_k that earns the highest vote.

$$V(L_k) = \frac{1}{M}\sum_{i}^{N_k}\sum_{j}^{M} v(p_j, c_i) \tag{6.1}$$

where N_k is the number of colours in k^{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^{th} pixel p_j with respect to colour c_i, such that it equals to 1 if c_i is the current colour of p_j and otherwise it is set to 0, as defined in Eqn. (6.2).

$$v(p_j, c_i) = \begin{cases} 1 & \text{if } p_j \text{ is coloured with } c_i \\ 0 & \text{otherwise} \end{cases} \tag{6.2}$$

However, in some exceptional cases, if there are two or more labels competing with (almost) similar vote totals, a middle label of the average colour 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 colour-scheme detection and annotation, each RGB clothing segment is converted to HSV, then reduced to a maximum of 64 indexed colours. A colour-scheme model is designed and used with the Euclidean distance matching method to categorize each colour into one of three colour-scheme categories ‘Warm’, ‘Cool’, and ‘Neutral’ as shown in Fig. 6.4. Since that, the three-valued colour 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 colours resulting in the same (squared average-sum) distance. Thus, all matched colours in the model and in the clothing segment are properly mapped to unique colour codes to avoid confusion and mismatch.

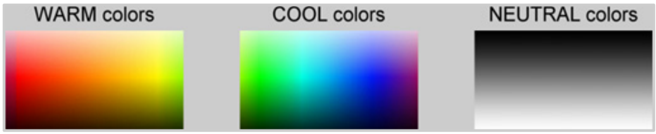


Figure 6.4: RGB colour representation for the colour-scheme model

Hence, a set of K (one to three) labelled colour groups is used with the same multi-case voting formulation described in Eqn. (6.1) and (6.2) to nominate the winning label L_k. Our colour-scheme model classifies all grayscale colours as 'Neutral'. Moreover, if two different colour-schemes (almost) equally exist in the upper or lower segment we also assign the label 'Neutral'. For the 'Overall colour-scheme' annotation we add the label 'Mixed', reflecting that upper and lower clothing have two different colour-schemes, as they are two separate (i.e. upper and lower) clothing pieces, obviously representing a new notable clothing style. Eventually the dominant colours for each of the upper and lower clothing segments are detected and semantically annotated. This is performed via a method for colour segmentation using a k-means clustering algorithm [112], which utilizes the information of colour histograms to select the dominant colours from the input image to be initialized as the k-means for the clustering process.

As such, each RGB clothing segment is converted to (four to nine) indexed colours 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 Eqn. (6.3), with two cases defined in Eqn. (6.4), is used to deduce the final integer value K' where $1 \leq K' \leq 3$ representing the number of estimated clusters (colours) in the image to be either one, two, or three, which implicitly describes the number of dominant colours 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 colours. Since the background, in a clothing segment, has the same colour 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/colour such that:

$$K' = \left(\sum\nolimits_{i}^{K} \omega(u_i)\right) - 1 \tag{6.3}$$

where

$$\omega(u_i) = \begin{cases} 1 & \text{if } \frac{|u_i|}{M} \geq t \\ 0 & \text{otherwise} \end{cases} \tag{6.4}$$

where ω is the weighting function of i^{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 t 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.

6.2.4 Analysing Patterns

A clothing segment is processed to extract and analyse 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 colour image before it is converted to grayscale. We use *uniform* Local Binary Pattern (ULBP) [113] as in Eqns. (6.5), (6.6), and (6.7). Here for detecting local binary patterns, for each 3×3 patch p we use a basic LBP operator defined with a circular symmetric neighbour set of *P*=8 members on a circle of radius equals to 1, as shown in Fig. 6.5. Essentially, the basic LBP operator produces for a 3×3 patch p an 8-bit binary number representing the detected local binary pattern such as 00001100, where each bit is assigned a value 0/1 by comparing the centre pixel g_c with each neighbour pixel (i.e. g_0 to g_7), as in Eqn. (6.5).

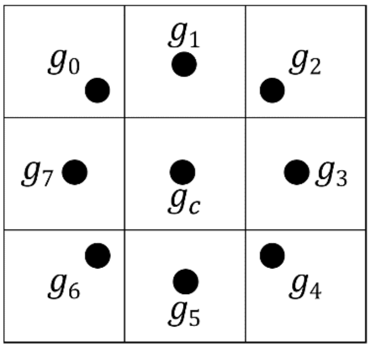


Figure 6.5: LBP operator with a circular symmetric neighbour set of *P*=8 members on circle of radius =1

Since we use ULBP, we are only concerned with rotation invariant uniform patterns that have a value of at most 2. The value 2 indicates that only zero, one, or two transitions are considered as uniform, where a transition is a single change from 0 to 1 or vice versa in a binary pattern. For instance, the pattern 11111111 (corresponding to 255) has no transitions so the number of transitions = 0 whereas the pattern 00001100 (corresponding to 12) has two transitions, which means that the number of transitions = 2; both values are deduced using Eqn. (6.7). Thus, the uniform is adopted to be wherever the number of transitions ≤ 2, namely adopting 58 values out of all possible integer grayscales, ranging from 0 to 255, where all the non-uniform patterns with the number of transitions > 2 are grouped as a miscellaneous pattern and eventually assigned the value 0.

$$s(x) = \begin{cases} 1 & \text{if } x > t \\ 0 & \text{if } x \leq t \end{cases} \tag{6.5}$$

where $s(x)$ is defined as the sign (either 0 or 1) of x denoting the difference between two grey values of the centre pixel g_c and a neighbour pixel g_i. Hence for each neighbouring grey value (g_i), a local neighbourhood is thresholded at the grey value of the centre pixel (g_c) into a binary pattern by computing the grey value difference $x = (g_i - g_c)$, to then set the 0/1 sign of difference $s(x)$, where $s(x)$ is set to 1 only if greater than a pre-defined threshold t (rather than 0 difference as in [113]). As such, we reduce the sensitivity of the $ULBP_P$ operator in Eqn. (6.6) by considering only the signs of grey value differences that are larger than t to be the minimum detectable gray-difference between two compared pixels g_i and g_c as stated in Eqn. (6.5). The resulting value of the function $f(p)$ represents the number of transitions of a local binary pattern in a 3×3 patch p being checked by Eqn. (6.7) whether it is one of uniform patterns if the value of $f(p) \leq 2$.

$$ULBP_P = \begin{cases} \sum_{i=0}^{P-1} s(g_i - g_c) & \text{if } f(p) \leq 2 \\ P + 1 & \text{otherwise} \end{cases} \tag{6.6}$$

where

$$f(p) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{i=1}^{P-1} |s(g_i - g_c) - s(g_{i-1} - g_c)| \tag{6.7}$$

Here the annotation procedure of clothing patterns includes the notion that a clothing pattern is more complex if it has more details and more colours 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 colour transitions between clothing and background, so edge points detected by Prewitt algorithm [114] are removed. Thus, we compute the pixel-sum by counting all non-zero pixels of the produced ULBP image. A prober label is assigned based on a pre-trained scale model defining the numeral range of the three labels ('None', 'Simple', and 'Complex').

6.3 Automatic 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 2.1, 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 of the scale.

Automatic comparative annotation is achieved via a major- and a sub-process as in Eqn. (6.8) and (6.9). 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 as L_i and L_j, which were earlier extracted and categorically annotated, as in Section 6.2.

$$compare_a(L_i, L_j) = \begin{cases} 1 & \text{if } L_i - L_j \leq -2 \\ 2 & \text{if } L_i - L_j = -1 \\ similarity_a(X_i, X_j) & \text{if } L_i - L_j = 0 \\ 4 & \text{if } L_i - L_j = 1 \\ 5 & \text{if } L_i - L_j \geq 2 \end{cases} \tag{6.8}$$

where

$$sim_a(X_i, X_j) = \begin{cases} 2 & \text{if } X_i < X_j \\ 3 & \text{if } X_i = X_j \\ 4 & \text{if } X_i > X_j \end{cases} \tag{6.9}$$

By using Eqn. (6.8), 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 sim_a is strong enough and to test if there is a slight difference that is undetectable by Eqn. (6.8). 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.

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 are derived using the same Ranking SVM method explained in Section 2.4, 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.

6.4 Automatic Soft Clothing Biometrics

6.4.1 Automatic Categorical Clothing Traits (*AutoCat-N*)

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. 6.9-a and Table 6.3-a), which are selected based on ANOVA (described in Section 4.3), 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*).

6.4.2 Automatic Comparative Clothing Traits (*AutoCmp-N*)

We use the ranking SVM method, explained in Section 2.4, 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 6.1. The coefficient *w* is derived by using the formulation in Eqn. (2.2). The desirable per attribute ordering of all subjects is derived by using *w*. Then by Eqn. (2.1), 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. 6.9-b and Table 6.3-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*).

6.5 Analysing Automatic Soft Clothing Data

6.5.1 Automatic Soft Clothing Description Data

Both front- and side-view samples of the Soton Gait subset (described in Section 2.2) are used for this experiment. Since we have already used this image data set in constructing the *manual* soft clothing biometric galleries (described in Section 2.5 and 5.2.2), we use it again here in conducting the automatic clothing attribute extraction and annotation, leading to construct new galleries of *automatic* soft clothing biometrics to be further used in subject identification and retrieval. This definitely enables the comparison between the manual and automatic soft clothing traits.

Table 6.2 shows a summary of all collected or inferred manual and automatic soft clothing annotations or comparisons. For the same subjects in database, soft body biometrics (described in 2.2.1) are used here again as a baseline for performance benchmarking and comparison with our proposed soft clothing biometrics and are also used as complementary traits supplemented by soft clothing traits.

Table 6.2: The obtained automatic/manual soft clothing data

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
***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

6.5.2 Correlations between Automatic Attributes

For better understanding of the usability and operability of the proposed soft clothing traits, the correlations between clothing attributes are investigated, interpreting how the attributes collaborate with each other to represent the biometric signature. The correlation matrices of categorical and comparative soft clothing traits (see Table 6.1) are shown in Fig. 6.6 and 6.7 respectively, where correlation coefficients are computed using Pearson's *r* and it is considered as significant when its corresponding p-value ≤ 0.05, as explained earlier in Section 3.1. With a view to clarifying and organizing the correlation analysis of attributes, the correlation matrix is sliced into three groups of attributes, as in Table 6.1. This not only enables the observation of the correlation between any two clothing attributes, but also highlights 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 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.

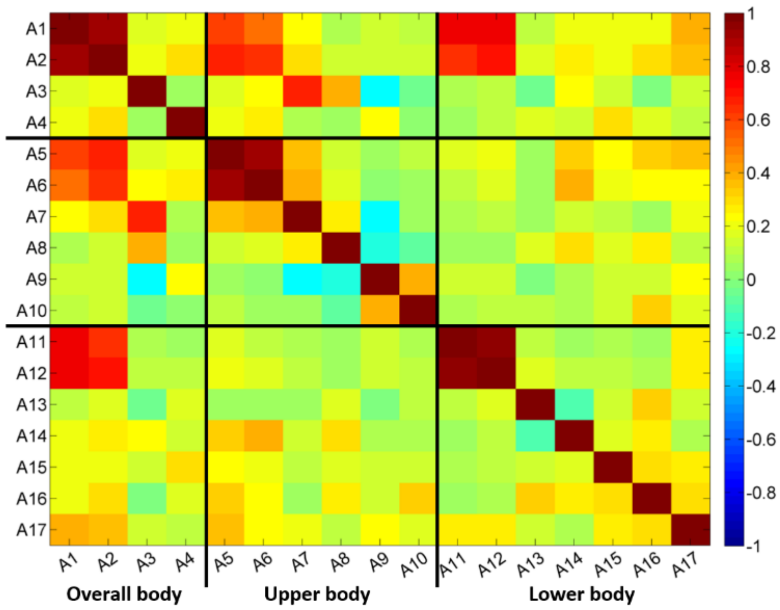


Figure 6.6: Correlations between automatic categorical soft clothing traits

As can be observed in Fig. 6.6, within all the three body parts – *overall*, *upper*, or *lower* – the relations between skin exposure 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 colour-scheme* (A8). This is consistent with the reality in that the *upper* (top) clothes worn by people have much more variety of colours and hues than the *lower* clothes, which in turn affect the *overall* appearance and style much more. Fig. 6.6 shows also that there are some attributes have multiple modest correlations with others, as follows: *Upper brightness* (A7) shows the same as *Lower colour-scheme* (A14) which is correlated with both (A5 and A6); *Upper pattern* (A10) with (A9 and A16); *Footwear category* (A17) correlated with (A1 and A2) as well; *Lower brightness* (A13) is correlated with *Lower pattern* (A16); and the attributes (A8 and A14), relating to colour-schemes, are correlated with each other.

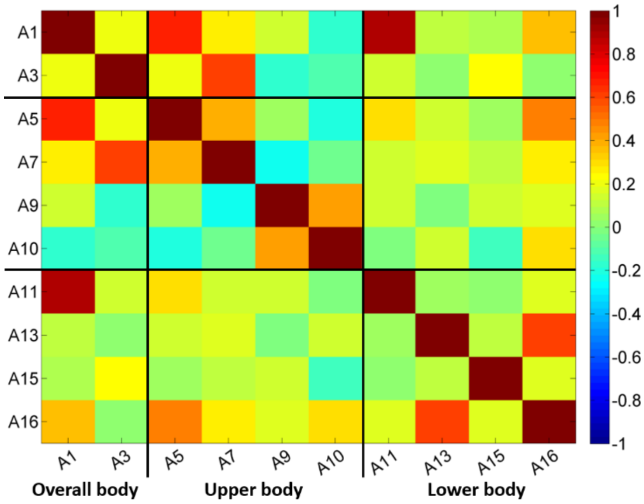


Figure 6.7: Correlations between automatic comparative soft clothing traits

The correlation matrix in Fig. 6.7, highlights a number of significant correlations between comparative traits including: the highest correlations of overall (A1) with (A5, A11) skin exposure traits; overall contrast (A3) is well correlated with upper brightness (A7); lower pattern (A16) is correlated with (A5) and also has slightly higher correlation with lower brightness (A13); and a good correlation between upper dominant colour (A9) and brightness (A10).

6.5.3 Distribution Statistics of Automatic Clothing Trait

Automatic soft clothing traits are analysed from statistical perspective to understand and compare their population in the clothing feature space, likewise the analysis applied to manual soft clothing traits in Section 3.2. Fig. 6.8 and 6.9 demonstrate the population of each categorical and comparative automatic traits and clarify the differences in their data distributions. It can be observed that a number of categorical and comparative traits have well-distributed data density along the available descriptive labels, such as overall contrast (A3), and upper brightness (A7). It is worth to reemphasize here the role of the influencing factors (discussed in Section 3.2), which affect the presence or absence and the dominancy or rarity of a particular trait in a database, will consequently affect the data distribution of that trait. For example, footwear category (A17) appears to be unbalanced distribution, since the greatest majority of people in database wearing close toed shoes and only few wearing open toed shoes, which might be a result of any cultural, environmental, seasonal etc. factors . Note that the relative trait data distributions, e.g.

clothing brightness and contrast (A3, A7, and A13) appear to be less affected by such factors.

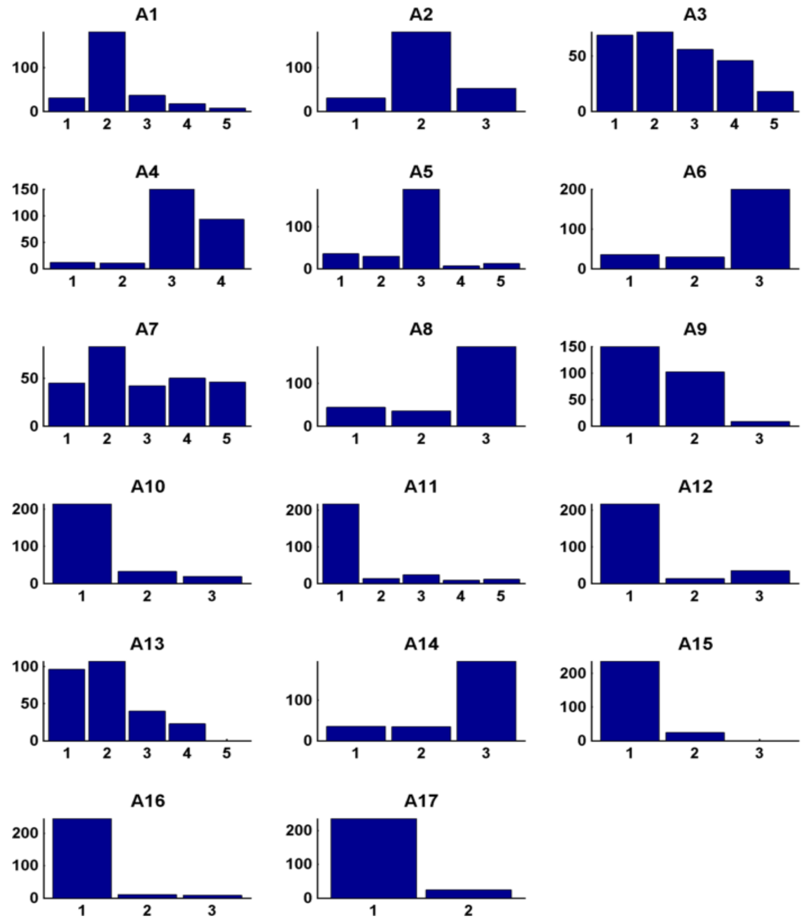


Figure 6.8: Data distribution of automatic categorical soft clothing traits

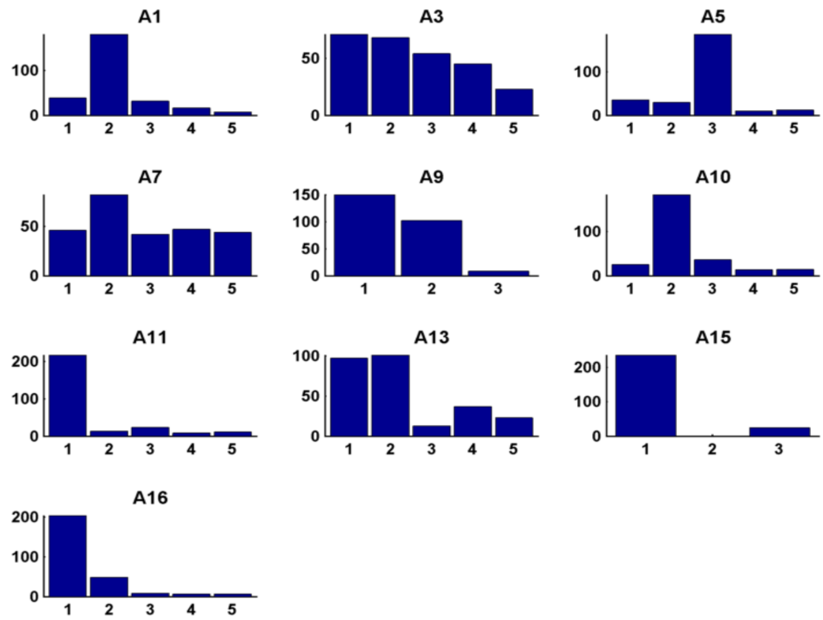
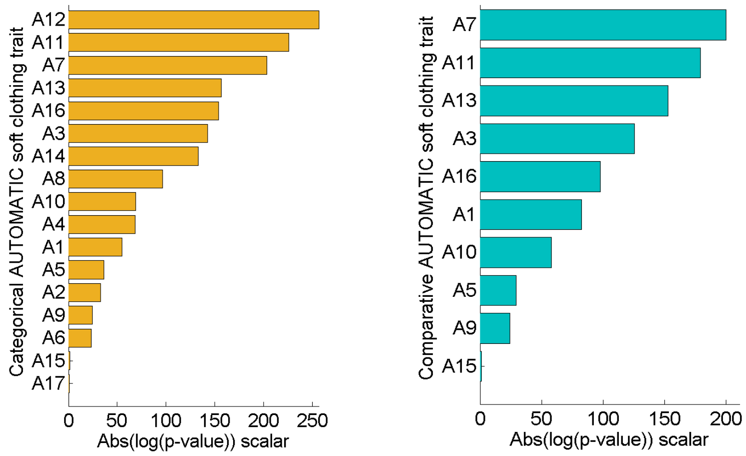


Figure 6.9: Data distribution of automatic comparative soft clothing traits

6.5.4 ANOVA of Automatic Clothing Traits

One-way ANOVA is used to determine the most effective and significant automatic soft clothing traits. The F-ratio and p-value are computed (using the method described in Section 3.3) per soft attribute for categorical and comparative traits, where the total 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 6.3 reports the resulting ANOVA test values for all clothing traits ordered by the best p-value. Accordingly, Fig. 6.10 depicts ordered p-values scaled positively by computing the absolute logarithm of the p-value, which 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 6.5.2. Like what we performed with manual soft clothing traits in Section 3.3, MANOVA is applied to the whole set of automatic soft clothing traits, exploring multivariate interactions between multiple attributes and highlighting significance of discrimination capability, as shown in Table 6.4, where all measures provide significant p-value <0.0001.



(a) Categorical clothing traits (b) Comparative clothing traits

Figure 6.10: Automatic soft clothing traits ordered according to p-values inferred by ANOVA

Table 6.3: ANOVA of automatic soft clothing traits ordered by p-value

(a) Categorical traits		
Clothing attribute	**F-ratio** ($df = 255$)	**p-value** ($p \leq 0.05$)
A12 Lower Clothing season	214.69	1.36E-111
A11 Lower Skin exposure	131.49	2.47E-98
A7 Upper Brightness	90.92	2.03E-88
A13 Lower Brightness	42.56	2.66E-68
A16 Lower Pattern	40.50	5.19E-67
A3 Upper vs Lower Contrast	33.59	3.73E-62
A14 Lower Colour-scheme	28.55	5.57E-58
A8 Upper Colour-scheme	15.18	2.96E-42
A10 Upper Pattern	9.03	2.97E-30
A4 Overall Colour-scheme	8.90	6.33E-30
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 colour	3.21	8.58E-11
A6 Upper Clothing season	3.12	2.33E-10
A15 Lower Dominant colour	0.82	8.62E-01
A17 Footwear Category	N/A	N/A

(b) Comparative traits		
Clothing attribute	**F-ratio** ($df = 255$)	**p-value** ($p \leq 0.05$)
A7 Upper Brightness	86.23	5.32E-87
A11 Low Skin exposure	61.46	5.58E-78
A13 Lower Brightness	39.78	1.52E-66
A3 Upper vs Lower Contrast	25.17	8.98E-55
A16 Lower Pattern	15.45	1.09E-42
A1 Overall Skin exposure	11.73	3.60E-36
A10 Upper Pattern	7.25	1.56E-25
A5 Up Skin exposure	3.69	6.07E-13
A9 Upper Dominant colour	3.21	8.57E-11
A15 Lower Dominant colour	0.82	0.86

Table 6.4: MANOVA of automatic soft clothing traits showing significance by different standard statistics

MANOVA statistic	**(a) Categorical traits**		**(b) Comparative traits**	
	F-ratio ($df = 255$)	**p-value** ($p \leq 0.05$)	**F-ratio** ($df = 255$)	**p-value** ($p \leq 0.05$)
Pillai's Trace	5.090	<0.0001	4.435	<0.0001
Wilks' Lambda	13.818	<0.0001	13.760	<0.0001
Hotelling's Trace	38.415	<0.0001	39.374	<0.0001
Roy's Largest Root	269.454	<0.0001	137.623	<0.0001

6.5.5 Mutual Dependence Analysis of Automatic Clothing Traits

- **Statistical Dependency**

Statistical dependency is computed for each automatic soft clothing trait using the same method described in Section 3.4.1. The resulting SD scores are used to list automatic traits in order as shown in Table 6.5. Upper brightness (A7) obtains the highest score followed by upper vs lower contrast (A3), in both categorical and comparative forms. Interestingly, categorical upper (A5) and then overall (A1) skin exposure offer much more dependency than lower exposure (A11), whereas comparative form of (A5) receives lower SD value compared with the comparative traits (A1) and (A11). In categorical and comparative lists, lower dominant colour (A15) consistently has the lowers SD value due to the large similarity across many people wearing similar clothing, especially trousers like jeans. In addition to the small variation in the number of dominant colour in lower clothes that mostly have a single dominant colour, which is not the case in upper clothes having much higher variety with regard to colour aspects.

Table 6.5: Statistical dependency (SD) analysis of automatic soft clothing traits

(a) Categorical traits	
Clothing attribute	**SD**
A7 Upper Brightness	3.905
A3 Upper vs Lower Contrast	3.729
A5 Upper Skin exposure	2.476
A1 Overall Skin exposure	2.427
A13 Lower Brightness	2.098
A11 Lower Skin exposure	2.088
A14 Lower Colour-scheme	1.922
A8 Upper Colour-scheme	1.894
A10 Upper Pattern	1.856
A9 Upper Dominant colour	1.657
A2 Overall Clothing season	1.633
A6 Upper Clothing season	1.394
A17 Footwear Category	1.189
A12 Lower Clothing season	1.157
A16 Lower Pattern	1.137
A4 Overall Colour-scheme	1.080
A15 Lower Dominant colour	0.647

(b) Comparative traits	
Clothing attribute	**SD**
A7 Upper Brightness	8.670
A3 Upper vs Lower Contrast	8.252
A13 Lower Brightness	8.148
A1 Overall Skin exposure	7.669
A16 Lower Pattern	7.597
A11 Low Skin exposure	7.561
A5 Upper Skin exposure	7.356
A10 Upper Pattern	7.136
A9 Upper Dominant colour	1.381
A15 Lower Dominant colour	0.687

- **Mutual Information**

Mutual information is measured for each automatic clothing trait using the same method described in Section 3.4.2, where the MI values are deduced from two different perspectives: (i) measuring MI between each trait and the subject IDs (representing the class labels), as presented in order in Table 6.6; and (ii) measuring MI between each two traits, as demonstrated in Fig. 6.11. Table 6.6 provides the same comparative trait ranking obtained by SD in Table 6.5. With respect to the categorical trait ranking by MI, the traits relating to brightness and contrast (A7, A3, A13) locate in the top of the rank as the most mutually informative automatic traits. In overview, each categorical upper clothing traits appears to be more mutually informative than its lower clothing rival, highlighting increased usefulness of information derived from upper body clothing for identity.

Table 6.6: Mutual information (MI) analysis of automatic soft clothing traits

(a) Categorical traits

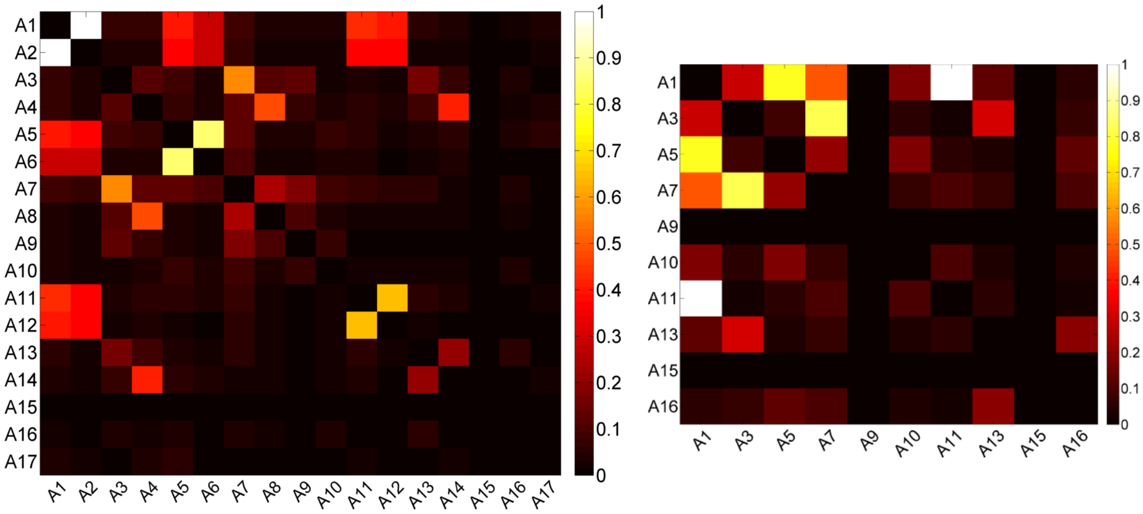
Clothing attribute	MI $I(X,Y)$
A7 Upper Brightness	2.178
A3 Upper vs Lower Contrast	2.029
A13 Lower Brightness	1.479
A1 Overall Skin exposure	1.088
A8 Upper Colour-scheme	1.054
A14 Lower Colour-scheme	0.990
A5 Upper Skin exposure	0.985
A2 Overall Clothing season	0.948
A9 Upper Dominant colour	0.824
A6 Upper Clothing season	0.763
A11 Lower Skin exposure	0.738
A10 Upper Pattern	0.692
A12 Lower Clothing season	0.616
A17 Footwear Category	0.404
A4 Overall Colour-scheme	0.287
A15 Lower Dominant colour	0.248
A16 Lower Pattern	0.239

(b) Comparative traits

Clothing attribute	MI $I(X,Y)$
A7 Upper Brightness	3.140
A3 Upper vs Lower Contrast	3.110
A13 Lower Brightness	3.091
A1 Overall Skin exposure	3.013
A16 Lower Pattern	2.990
A11 Low Skin exposure	2.982
A5 Up Skin exposure	2.960
A10 Upper Pattern	2.882
A9 Upper Dominant colour	0.714
A15 Lower Dominant colour	0.038

Fig. 6.11 shows normalized MI measurements (ranging from 0 to 1) of all possible pairs of automatic soft clothing traits, where high MI values symbolised by (white, yellow), medium values by (orange, red), and low values by (brown, black). All diagonal cells are set to zero, to be shown in (black) as irrelevant quantities. In the categorical MI matrix, each of the three skin exposure traits (A1, A5, A11) is highly mutually informative, when paired with its corresponding clothing season trait (A2, A6, A12) respectively. Due to the high variety of upper clothes hues and colours compared to lower clothes, upper vs lower contrast (A3) shows a large MI value with upper brightness (A7) but not with lower brightness (A13). Overall skin exposure (A1) and season (A2), consistently display good mutual information with the related traits (A5, A6, A11, A12). Overall colour-scheme (A4) offers good MI values with upper (A8) and lower (A14) colour-schemes. In the

comparative matrix, the most significant mutual information exists between the skin exposure traits (A1, A5, A11), also between brightness and contrast traits (A7, A13, A3).



(a) Categorical clothing traits (b) Comparative clothing traits

Figure 6.11: Mutual information (MI) matrix of soft clothing traits

6.6 Viewpoint Invariant Retrieval by Automatic Soft Clothing Traits

We conduct a number of experiments to achieve soft biometric-based retrieval and we adopt the same challenging methodology used in Section 5.2.3 to evaluate the use of soft clothing biometrics for front/side viewpoint invariant retrieval. For all soft biometric galleries described in Table 6.7, all feature vectors derived by describing front-view subject images are kept in their 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 [27]. 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 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 Colour);
- supplementing soft body biometrics to achieve enhanced retrieval by soft clothing traits added to a subset of 17 invariant soft body traits ; and
- achieving successful retrieval using soft clothing traits alone.

Soft body descriptions are available for the current Soton dataset, which were previously collected [39], such that each of 128 subjects was labelled 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 Colour*). 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 us to examine and evaluate the capability and performance of soft clothing biometrics in subject retrieval. Table 6.7 describes the ensuing biometric galleries of manual/automatic 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 neighbour 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: CMC; ROC with its relevant measurements AUC, EER; the Decidability Index (d'); and the F-measure known as (F1 score) deduced from associated $precision$ and $recall$ metrics, which is computed by averaging all computed F-measure values along multiple thresholds, where F1$= 2\, precision/recall$. Note that the larger the F1, the better is the performance. With respect to all metrics, all approaches are ranked according to their overall performance. More details about performance evaluation measurements can be found in Section 4.1.1.

Table 6.7: Automatic/manual clothing and body galleries for retrieval testing and comparison

	Clothing-based soft biometrics	Description
MANUAL	***Cat-6***	6 *manual* categorical clothing traits; the best correlated and most discriminative via ANOVA
MANUAL	***Cmp***	7 *manual* comparative soft clothing traits
AOTOMATIC	***AutoCat-14***	Top 14 *automatic* categorical clothing traits via ANOVA
AOTOMATIC	***AutoCmp-5***	Top 5 *automatic* comparative clothing traits via ANOVA
AOTOMATIC	***Auto(Cat-14&Cmp-5)***	Fusion of ***AutoCat-14*** and ***AutoCmp-5***

Body-based soft biometrics	Description
tradSoft	4 categorical soft body biometrics (Age, Ethnicity, Sex, and Skin Colour)
softBody	17 categorical soft body biometrics including ***tradSoft*** (observable from front/side views)

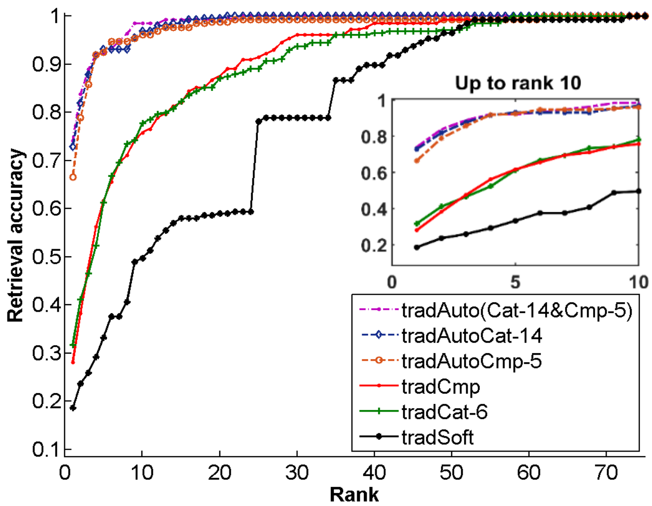
	Combined soft clothing and body biometrics	Description
Clothing & *tradSoft*	***tradAutoCat-14***	***AutoCat*-14** combined with ***tradSoft***
Clothing & *tradSoft*	***tradAutoCmp-5***	***AutoCmp-5*** combined with ***tradSoft***
Clothing & *tradSoft*	***tradAuto(Cat-14&Cmp-5)***	***Auto(Cat-14&Cmp-5)*** combined with ***tradSoft***
Clothing & *tradSoft*	***tradCat-6***	***Cat-6*** combined with ***tradSoft***
Clothing & *tradSoft*	***tradCmp***	***Cmp*** combined with ***tradSoft***
Clothing & *softBody*	***softAutoCat-14***	***AutoCat*-14** combined with ***softBody***
Clothing & *softBody*	***softAutoCmp-5***	***AutoCmp-5*** combined with ***softBody***
Clothing & *softBody*	***softAuto(Cat-14&Cmp-5)***	***Auto(Cat-14&Cmp-5)*** combined with ***softBody***
Clothing & *softBody*	***softCat-6***	***Cat-6*** combined with ***softBody***
Clothing & *softBody*	***softCmp***	***Cmp*** combined with ***softBody***

6.6.1 Retrieval by Automatic Clothing Added to Traditional Soft Biometrics

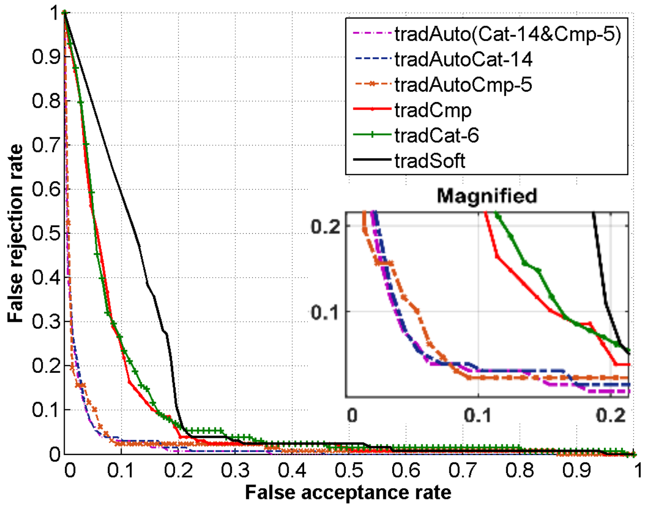
Soft clothing biometrics are used to supplement the traditional soft biometrics (*Age*, *Ethnicity*, *Sex*, and *Skin Colour*) to achieve enhanced person retrieval, since these four traits are commonly used in the literature and considered as essential (global) soft traits to be often embedded with other body or face soft traits, so here we embed them with clothing soft traits. Retrieval performance scores are reported in Table 6.8 for the use of traditional soft traits in isolation (i.e. *tradSoft*) in comparison with the five supplemented by clothing approaches (described in Table 6.7). Fig. 6.12-a, -b, and -c show performance curves in terms of CMC (Fig. 6.12-a), ROC (Fig. 6.12-b), and Precision-recall (Fig. 6.12-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. 6.12-a, -b, and -c, and as consistently suggested by the overall rank in Table 6.8. The *tradAuto(Cat-14&Cmp-5)* approach outperforms all its counterparts in all 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. 6.12-b) and Precision-recall (Fig. 6.12-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*. The average F-measure results and the corresponding curves in Fig. 6.12-c indicate that there is likely to be a large number of false positives (acceptances) against much fewer false negatives (rejections), resulting in high recall rates and lower precision ratios. Fig. 6.13 presents genuine and imposter distributions along with their consequent acceptance and rejection error rates, which clarifies and compares the matching performance of the examined clothing approaches.

Table 6.8: CMC and ROC metrics of retrieval using soft clothing traits with traditional soft biometrics

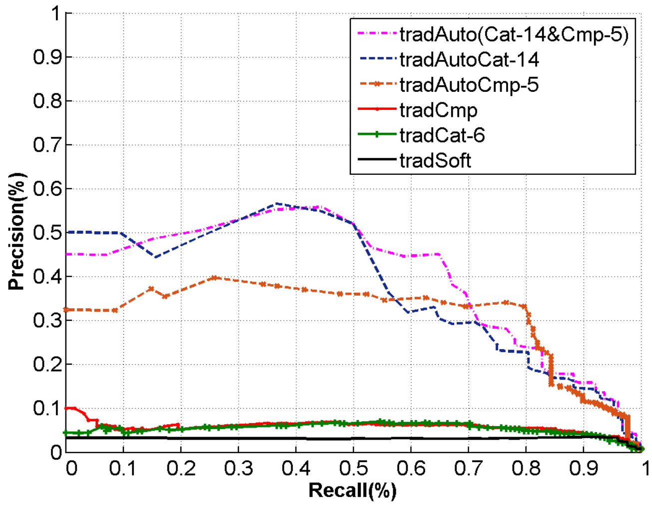
Approach	Top rank	AVG sum match scores up to rank		100% accuracy achieved at rank	EER	AUC	*d'*	Avg. F1	Overall rank
	=1	*=10*	*=128*						
tradSoft	0.187	0.345	0.874	73	0.203	0.129	1.857	0.331	6
tradCat-6	0.318	0.595	0.940	58	0.145	0.089	1.249	0.421	5
tradCmp	0.281	0.588	0.944	69	0.130	0.081	1.975	0.427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



(a) CMC performance up to rank 75

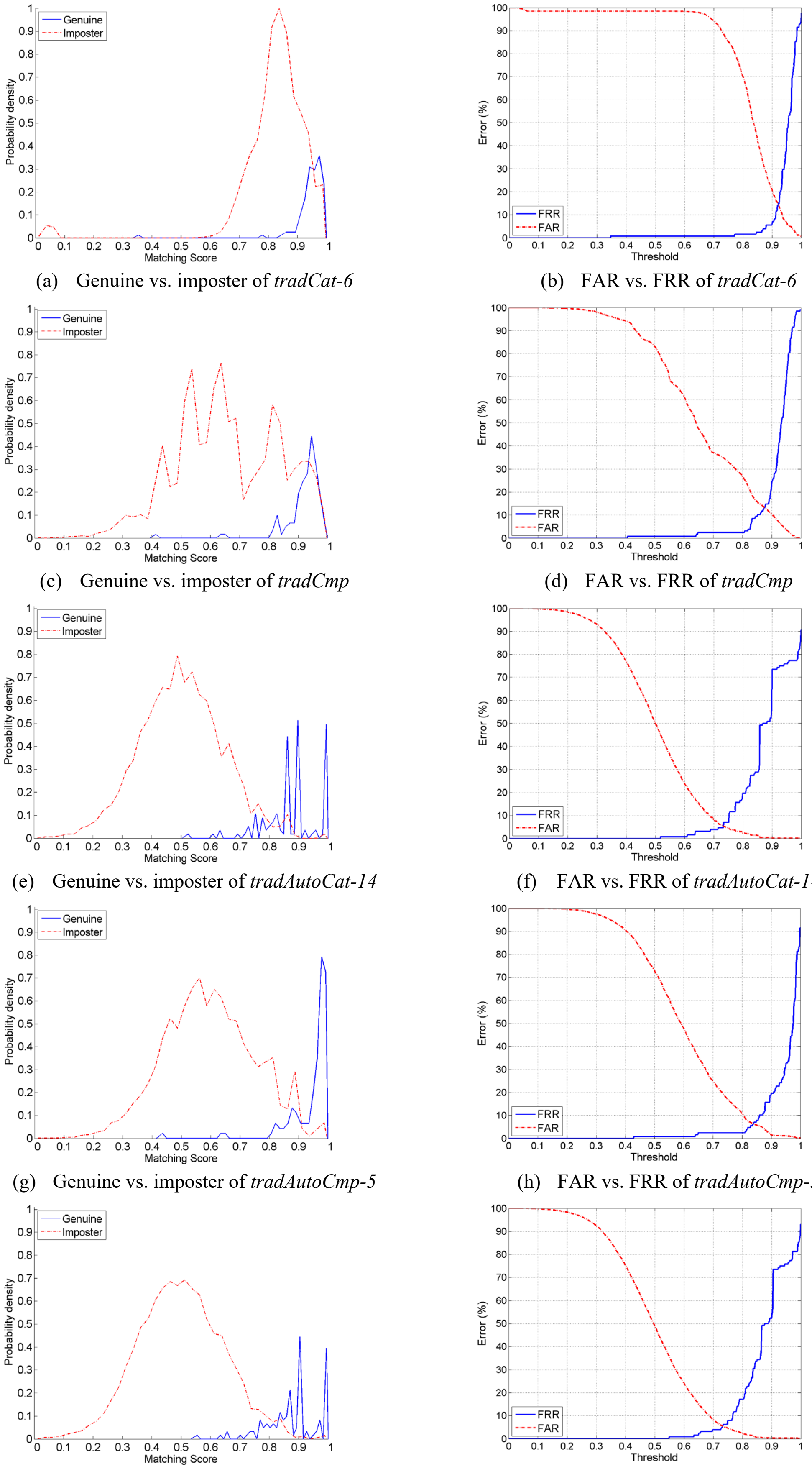


(b) ROC curves



(c) Precision-recall curves

Figure 6.12: CMC and ROC performance of soft clothing traits added to traditional soft biometrics



(a) Genuine vs. imposter of *tradCat-6*

(b) FAR vs. FRR of *tradCat-6*

(c) Genuine vs. imposter of *tradCmp*

(d) FAR vs. FRR of *tradCmp*

(e) Genuine vs. imposter of *tradAutoCat-14*

(f) FAR vs. FRR of *tradAutoCat-14*

(g) Genuine vs. imposter of *tradAutoCmp-5*

(h) FAR vs. FRR of *tradAutoCmp-5*

(i) Genuine vs. imposter of *tradAuto(Cat-14&Cmp-5)*

(j) FAR vs. FRR of *tradAuto(Cat-14&Cmp-5)*

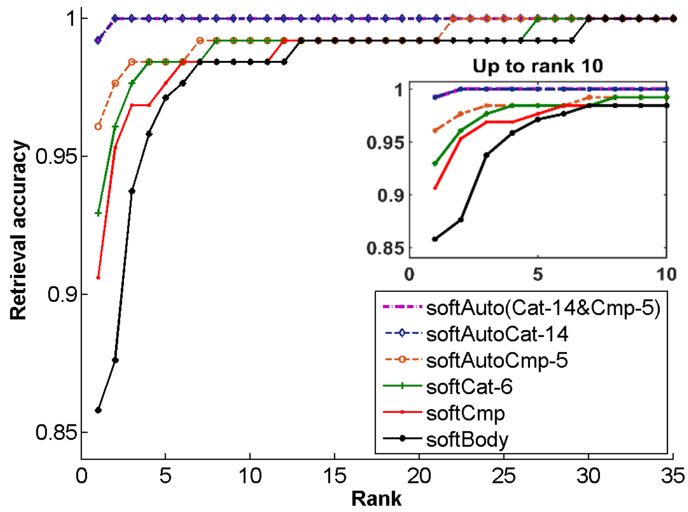
Figure 6.13: Genuine vs imposter and error curves of automatic clothing traits with *tradSoft* for retrieval

6.6.2 Retrieval by Automatic Clothing Added to Soft body Biometrics

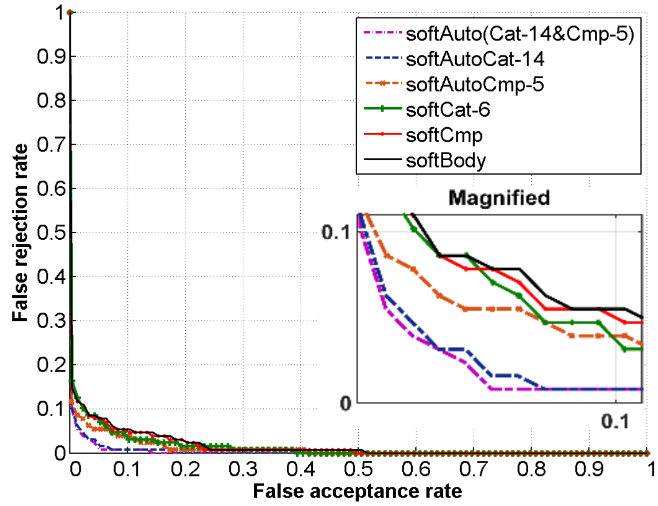
Soft body biometrics are supplemented using soft clothing biometrics for enhanced person retrieval, targeting the highest possible performance via the full power of soft body traits and the efficacy of soft clothing traits. CMC, ROC, and Precision-recall curves are provided in Fig. 6.14-a, -b, and -c 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 6.7). Moreover, the consequent CMC scores and the ROC analysis of all approaches are illustrated in Table 6.9. 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 6.9 and performance curves in Fig. 6.14-a, -b, and -c. Besides, Fig. 6.15 shows and compares genuine and imposter densities besides FAR and FRR curves. Soft clothing effectively supplements soft body retrieval resulting in considerable improvement in both precision and recall ratios, as can be observed in Fig. 6.14-c.

Table 6.9: CMC scores and ROC metrics of retrieval using soft clothing traits with soft body biometrics

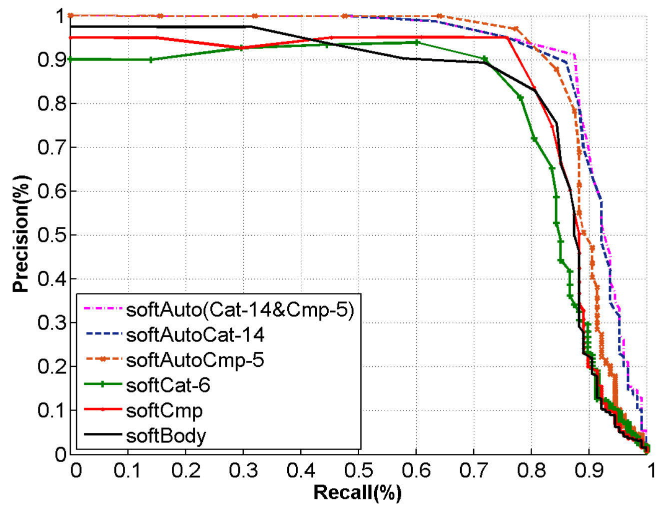
Approach	**Top rank**	**AVG sum match scores up to rank**		**100% accuracy achieved at rank**	**EER**	**AUC**	***d'***	**Avg. F1**	**Overall rank**
	=1	***=10***	***=128***						
softBody	0.858	0.952	0.995	30	0.068	0.016	3.372	0.638	6
softCat-6	0.930	0.978	0.997	27	0.063	0.013	3.082	0.639	4
softCmp	0.906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



(a) CMC performance up to rank 35

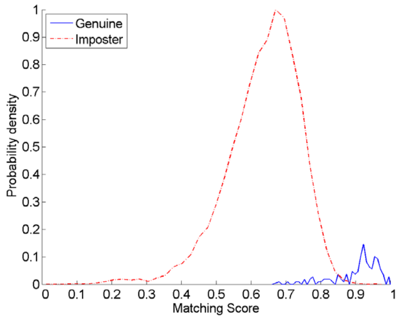


(b) ROC curves

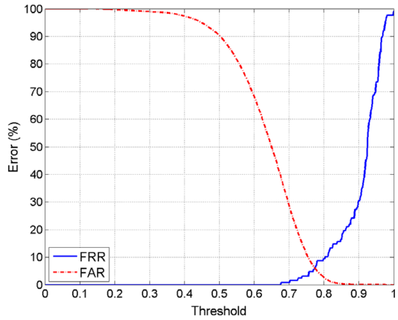


(c) Precision-recall curves

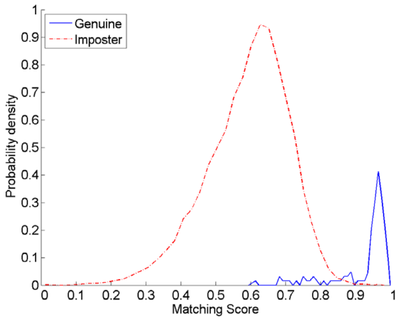
Figure 6.14: CMC and ROC performance of soft clothing traits added to soft body biometrics



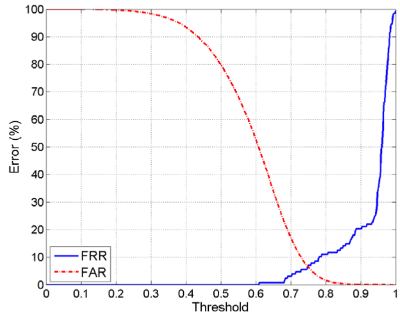
(a) Genuine vs. imposter of *softCat-6*



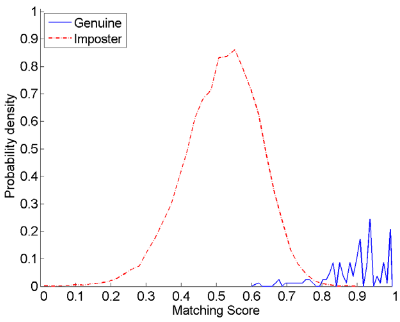
(b) FAR vs. FRR of *softCat-6*



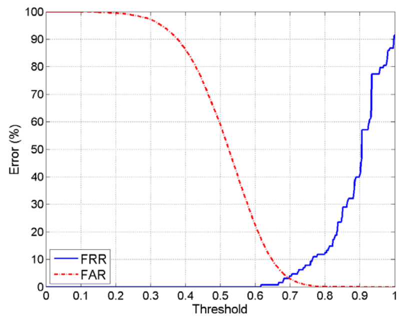
(c) Genuine vs. imposter of *softCmp*



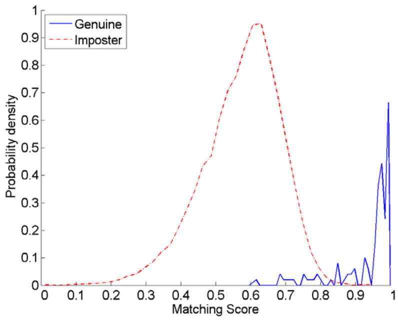
(d) FAR vs. FRR of *softCmp*



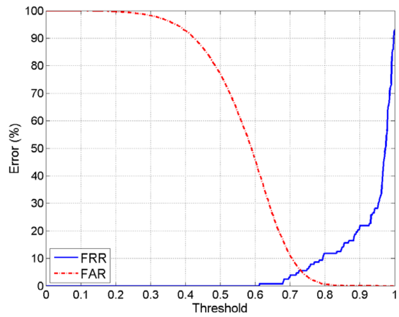
(e) Genuine vs. imposter of *softAutoCat-14*



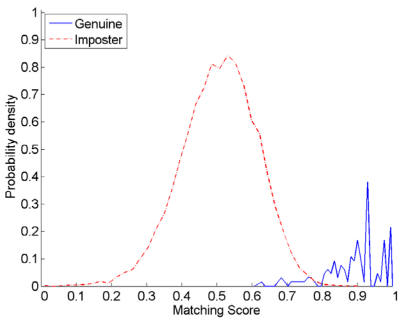
(f) FAR vs. FRR of *softAutoCat-14*



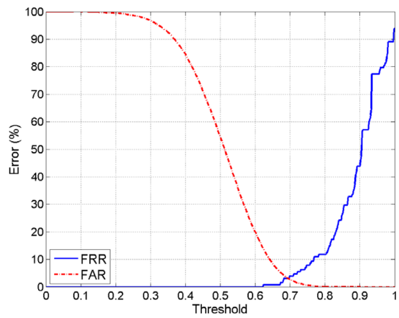
(g) Genuine vs. imposter of *softAutoCmp-5*



(h) FAR vs. FRR of *softAutoCmp-5*



(i) Genuine vs. imposter of *softAuto(Cat-14&Cmp-5)*



(j) FAR vs. FRR of *softAuto(Cat-14&Cmp-5)*

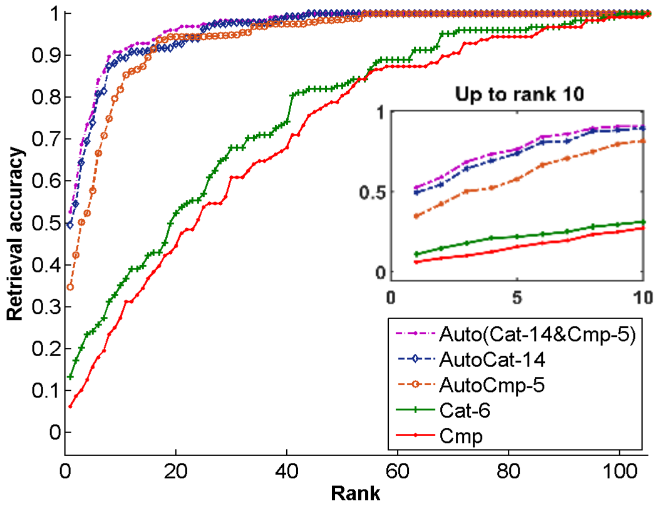
Figure 6.15: Genuine vs imposter and error curves of automatic clothing traits with *softBody* for retrieval

6.6.3 Retrieval using Automatic Clothing Biometrics Alone

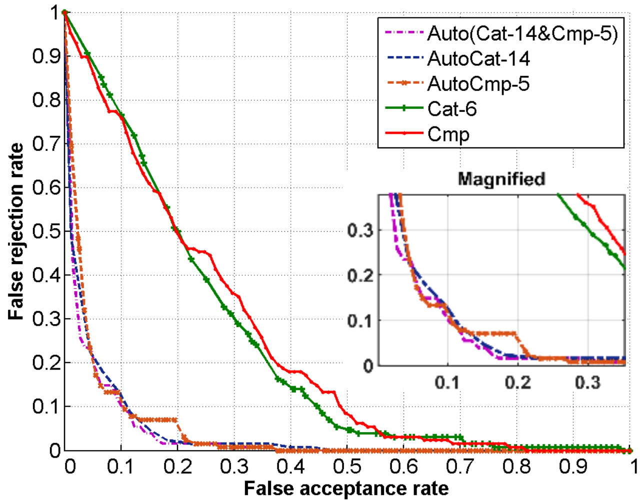
The clothing-based soft traits are used in isolation, aiming to achieve successful retrieval and to investigate the max potency of pure soft clothing traits when used alone in person retrieval, which is deemed to be a mimic of challenging real-case forensic scenarios when soft clothing attributes are the only observable soft traits. The retrieval evaluation results and performance analysis for all five clothing-based approaches (described in Table 6.7), are illustrated in Table 6.10 along with Fig. 6.16-a, -b, and -c. 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 in 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 performance. This is reflected by all associated curves in Fig. 6.16 and metrics in Table 6.10. 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. 6.16-b. It also achieves a better average F1 score, considering the trade-off between the estimated precision and recall represented in Fig. 6.16-c. *Cat-6* also surpasses *Cmp* here in all CMC and ROC metrics, but $\boldsymbol{d'}$ and F1, which are observed to be better and higher by *Cmp*. In all aspects when soft clothing are used alone, 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, which likely decreases the precision ratios of all approaches, even though they still achieve high recall ratios, as can be observed in Fig. 6.16-c. Fig. 6.17 shows genuine and imposter distributions and highlights the trade-offs between FAR and FRR curves of each approach, which confirms the potency of the automatic soft clothing traits in subject retrieval when used alone, as when used in fusion.

Table 6.10: CMC scores and ROC metrics of retrieval using soft clothing traits alone

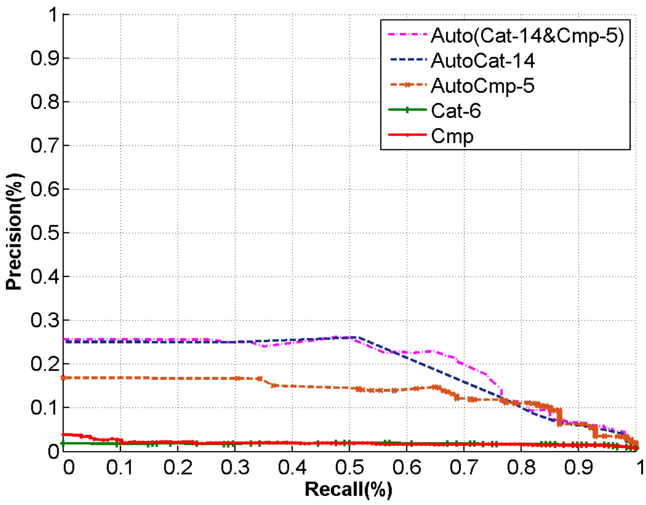
Approach	**Top rank**	**AVG sum match scores up to rank**		**100% accuracy achieved at rank**	**EER**	**AUC**	***d'***	**Avg. F1**	**Overall rank**
	=1	***=10***	***=128***						
Cat-6	0.133	0.224	0.806	100	0.303	0.232	0.630	0.250	4
Cmp	0.063	0.166	0.773	105	0.319	0.240	1.056	0.275	5
AutoCat-14	0.495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



(a) CMC performance up to rank 105

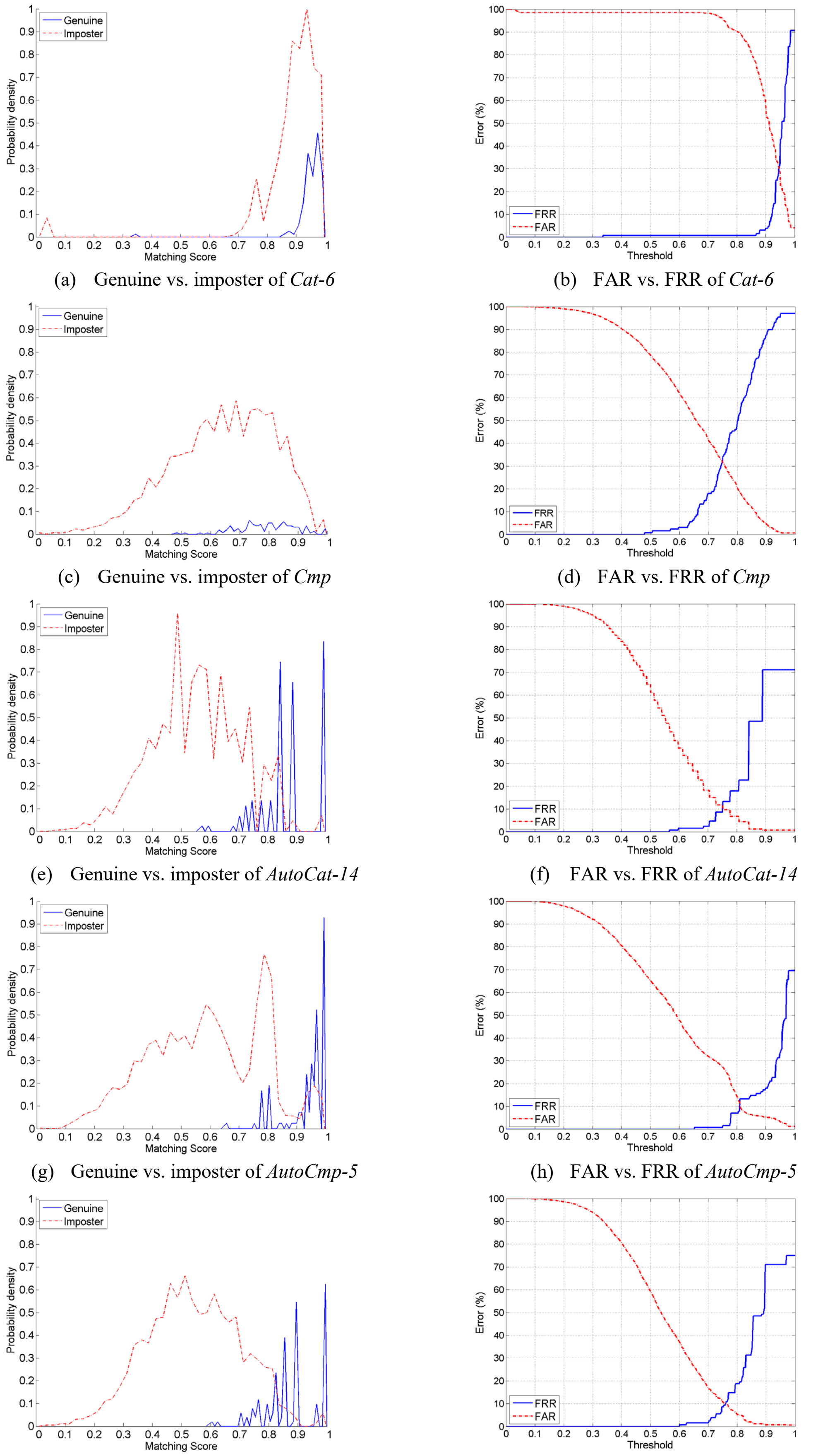


(b) ROC curves



(c) Precision-recall curves

Figure 6.16: CMC and ROC performance of soft clothing traits alone



(a) Genuine vs. imposter of *Cat-6*

(b) FAR vs. FRR of *Cat-6*

(c) Genuine vs. imposter of *Cmp*

(d) FAR vs. FRR of *Cmp*

(e) Genuine vs. imposter of *AutoCat-14*

(f) FAR vs. FRR of *AutoCat-14*

(g) Genuine vs. imposter of *AutoCmp-5*

(h) FAR vs. FRR of *AutoCmp-5*

(i) Genuine vs. imposter of *Auto(Cat-14&Cmp-5)*

(j) FAR vs. FRR of *Auto(Cat-14&Cmp-5)*

Figure 6.17: Genuine vs imposter and error curves of automatic clothing traits *alone* for subject retrieval

6.7 Conclusions

In this chapter a set of proposed automatic soft clothing attributes is introduced and briefly described. The methodology used for vision-based automatic extraction and annotation is explored. The ensuing automatic soft clothing biometrics and the detailed analysis of these traits are presented. A number of experiments on viewpoint invariant retrieval are demonstrated, investigating the power of automatic soft clothing traits derived via computer-vision and allowing performance comparison against those manual soft clothing traits inferred by humans.

We show that it is possible to 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. Automatic 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 automatic soft clothing traits are shown to provide more objective information and to offer increased discriminatory capabilities than manual soft clothing traits. As what we found out earlier about the manual approaches, the automatic approaches are shown to be relatively immune to change from the front to the side viewpoint. The high scoring retrieval received by soft clothing biometrics and the success in retrieving a subject-of-interest within a relatively small list (e.g. 10 subjects), indicate a promising potential of effective use in various biometric applications. In this way it narrows the search to a much smaller number of candidates. Although the subject-of-interest may not always be the top match in retrieval, when attempting to answer the question "Is the top match correct?", retrieval could still answer an alternative question: "Is the correct answer in the top *k* matches?". 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 are both describing the same relative attribute, they offer different descriptions and varied information on identity in a way that is mutually supplemental.

The obtained experimental results suggest that using the automatic comparative version of the relative soft clothing traits yields improved performances with increased discrimination over using the categorical version of the same traits. In general, relative soft clothing traits appear to be more significant in composing a discriminative biometric signature of the subject. This makes them more likely to be consistently described along all images in the database and can be similarly applied further to describe a new subject not in the dataset.

Though, automatic soft clothing traits are used –like manual soft clothing traits– in a manner to bridge the semantic gap between machine and human, they are intended further to bridge another semantic gap between vision-based low-level features and high-level clothing annotations. As such, it is still desirable for automatically assigned attribute annotations to be logically and generally acceptable by human observation and judgment, assuring the correct and complete bridging of both semantic gaps. Since this initial study aims to convey an important message that clothing information can be differentiated as soft biometrics, as part of a program aimed to achieve the capability to automatically search video and images in the wild for identification purposes.

Chapter 7

On the Capabilities of Soft Biometrics using Clothing Attributes

7.1 Evaluating Clothing Feature Performance

In this analysis, we utilize five different methods to evaluate then rank or list in order the manual/automatic soft clothing features by either their ranking scores or suggested feature selection ordering. This analysis leads to better empirical understanding of how significant and relevant are those clothing features to the human identification problem. Moreover, the investigation of feature significance and relevance can be used further to guide the feature subset selection, achieving the highest possible performance with a minimum number of features, as will be explored in Section 7.2.

This analysis is applied to the main four soft clothing forms as follows:

- Manual categorical soft clothing traits referred to as (*ManCat-21*).
- Manual comparative soft clothing traits referred to as (*ManCmp-7*).
- Automatic categorical soft clothing traits referred to as (*AutoCat-17*).
- Automatic comparative soft clothing traits referred to as (*AutoCmp-10*).

For this analysis, the adopted scoring methods comprises: ANOVA, SD, and MI which are explained in Section 3.3, 3.4, 6.5.4, and 6.5.5, where the resulting scores are used to rank clothing traits per each method. In addition, two more feature subset selection methods are used to order features by their cooperative usefulness in recognition based on the way of ordering (or prioritising in selection) enforced by inclusion/exclusion steps per each method. These two methods are Sequential Forward Selection (SFS) and Sequential Floating Forward Selection (SFFS).

Throughout this chapter, 'M' denotes the type of *manual* traits (introduced in Chapter 2) to differentiate them from the *automatic* traits 'A' (presented in Chapter 6). Besides, all experiments for clothing trait performance evaluation are achieved using the challenging methodology of viewpoint invariant retrieval described in Section 5.2.3.

Recognition rates over the feature subset size increase are depicted for all approaches at three ranks: rank 1 (the best match at the top rank); rank 10 (the capability of correct retrieval within a list of top ten matches); and rank 128 (the average accuracy over all matches at the full-rank). Standard Error of the Mean (SEM) associated with estimating an average-score are shown for rank 10 and rank 128 and computed as $\pm\mathrm{SEM} = \sigma/\sqrt{k}$ where σ is the standard deviation of all scores up to rank k.

In the figures 7.1, 7.5, 7.9, and 7.13, which show and compare the ranking scores of all traits ordered by the average of normalized scores for *ManCat-21*, *ManCmp-7*, *AutoCat-17*, and *AutoCmp-10* respectively. The corresponding scores of all traits are normalized per each scoring method (i.e. ANOVA, SD, and MI) using min-max normalization to rescale all values into the range [0, 1]. As such, given $\mathbb{S}$ as a set of all scores inferred by a single method, for each score $s \in \mathbb{S}$, a new normalized score is deduced as $s' = (s - min_{\mathbb{S}})/(max_{\mathbb{S}} - min_{\mathbb{S}})$, where $min_{\mathbb{S}}$ is the minimum score in $\mathbb{S}$ and $max_{\mathbb{S}}$ is the maximum score in $\mathbb{S}$.

- **SFFS vs. SFS**

SFS and SFFS are well known and widely used feature subset selection methods in practice [94]. They both work in similar way by starting with an empty set then iteratively include the next best trait to the subset of selection until stopping when the highest possible performance is reached.

The main difference is that SFFS is a (wrapper) method that performs some possible backward exclusion steps (of the worst traits) after each forward inclusion step (of the best trait), as long as the new subset (after the exclusion) increases the previous performance of a subset with the same size. This is what is known as “floating exclusion” step, which is applied by the Sequential Backward Selection (SBS) algorithm.

For more clarity and to allow comparison, SFS and SFFS algorithms can be given as follows:

SFS algorithm:

1. Start with the empty set
   $Y_0 = \emptyset$
2. Select the next best feature
   $x^+ = \arg\max_{x \notin Y_k} J(Y_k + x)$
3. Add the best feature to the set
   $Y_{k+1} = Y_k + x^+; \;\; k = k + 1$
4. Go to step 2

SFFS algorithm:

1. Start with the empty set
   $Y_0 = \emptyset$
2. Select the best feature
   $x^+ = \arg\max_{x \notin Y_k} J(Y_k + x)$
   $Y_{k+1} = Y_k + x^+; \;\; k = k + 1$
3. Select the worst feature
   (*consider book-keeping, avoiding infinite loops*)
   $x^- = \arg\max_{x \in Y_k} J(Y_k - x)$
4. If $J(Y_k - x^-) > J(Y_k)$ then
   $Y_{k-1} = Y_k - x^-; \;\; k = k - 1$
   Go to step 3
   Else
   Go to step 2

Unlike ANOVA, SD, and IM scoring methods, SFS and SFFS do not enforce ranking or scoring for each trait independently. It is rather that they enforce prioritising (in incrementally selecting) each trait in relation to the other traits in the selected subset. We utilize SFS and SFFS algorithms further for sorting traits based on a sequence suggested by either algorithm, in which a trait is the best as a next-selection to be added to a (cumulative) subset achieving the highest (incremental) recognition score. As such, with respect to the ordering suggested by SFFS algorithm (see Table 7.1), a sequence of traits comprising {M20, M11, M8} does not necessarily mean that M20 is the best among these three traits when used alone, nor that M11 is worse than M20 or better than M8 in performance. It can rather mean that a combination of {M20, M11} achieves better performance than a combination of {M11, M8} and a further combination of all {M20, M11, M8} is better than a combination of any two of them. This example is applicable on any three sequential traits suggested by SFS such as {M5, M20, M11}, in Table 7.1, except that only M5 is indeed the best trait in performance among all traits when used alone (i.e. feature subset size=1), since the first step of SFS algorithm tests each trait individually to find out which feature achieves the highest performance to be included in the feature subset and it is impossible to be excluded as in SFFS algorithm, where all selections even the first may be excluded from the feature subset for one or more times.

Note that, to obtain the SFS and SFFS ordering lists shown in the tables Table 7.1, 7.2, 7.3, and 7.4, we proceed to include all traits rather than terminating when only the best performance is achieved with reduced dimensionality. This is because we aim to prioritise all traits –not just a subset of the best trait as in the standard use– but even the ones may do not change or lower the performance, and to suggest ordering for the incremental feature subset selection starting from one trait and ending with the complete set of traits.

7.1.1 Evaluating Manual Categorical Clothing Features

Table 7.1 shows the ordering lists of manual categorical soft clothing traits (*ManCat-21*), which are deduced using SFS and SFFS methods. It can be observed that each method suggests a different ordering for the examined traits. Upper clothing category (M5) is reported earlier (in Section 3.3 and 3.4) to be the best trait by SD and MI, while it is not found as the best discriminative trait by ANOVA. M5 is also listed on top of SFS ordering, whereas it is dropped down the ordering list suggested by SFFS in Table 7.1. On the other hand, Lower body category (M9) is granted rank 1 only once by ANOVA and Style category (M20) is suggested as the first trait in SFFS ordering. In overview, Sleeve length (M8) appears to be the most useful trait, since it continuously remains at either rank 3 or 4 with respect to all rankings, also it is the third in SFFS ordering and the fourth in SFS ordering. Leg length (M11) is generally the next in usefulness, which appears within the range from 2 to 7 throughout all rankings and ordering lists. Fig. 7.1 shows and compares the ranking scores of *ManCat-21* traits ordered by the average of min-max normalized scores per method to the range [0, 1]. It can be seen that the four top traits (M8, M11, M20 and M6) indeed receive far higher scores by all ranking methods.

Table 7.1: Ordering manual categorical clothing traits using SFS and SFFS

***ManCat-21* Soft Clothing Traits**		
Ordering	**SFS**	**SFFS**
1	M5	M20
2	M20	M11
3	M11	M8
4	M8	M10
5	M10	M13
6	M13	M6
7	M6	M12
8	M7	M14
9	M1	M9
10	M12	M2
11	M15	M3
12	M14	M4
13	M9	M16
14	M2	M19
15	M3	M17
16	M4	M15
17	M16	M18
18	M18	M21
19	M19	M5
20	M21	M1
21	M17	M7

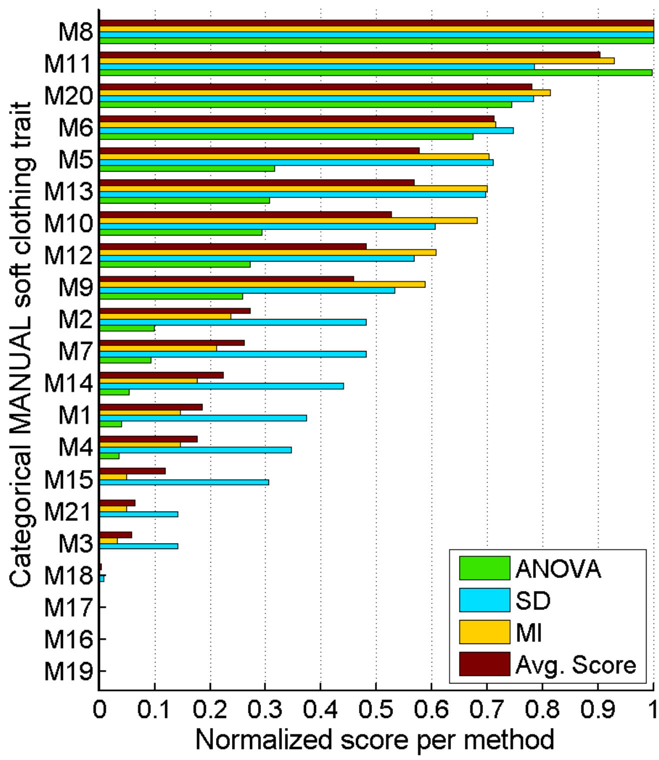


Figure 7.1: *ManCat-21* traits ordered by average min-max normalized ranking scores to the range [0, 1]

Fig. 7.2, 7.3, and 7.4 show recognition rates at rank 1, 10, and 128 respectively using *ManCat-21* clothing traits, which enable performance comparison between the five ranking or ordering methods along all possible sizes of the a feature subset (i.e. from 1 to 21). SFS attains the best performance on average and ANOVA reaches the highest possible performance using only 11 traits, as shown in Fig. 7.2. MI achieves the highest average performance overall subset sizes from one to 21 as in Fig. 7.3 and 7.4, while here again ANOVA obtains the best possible recognition rate at both 10 and 128 ranks using only ten traits.

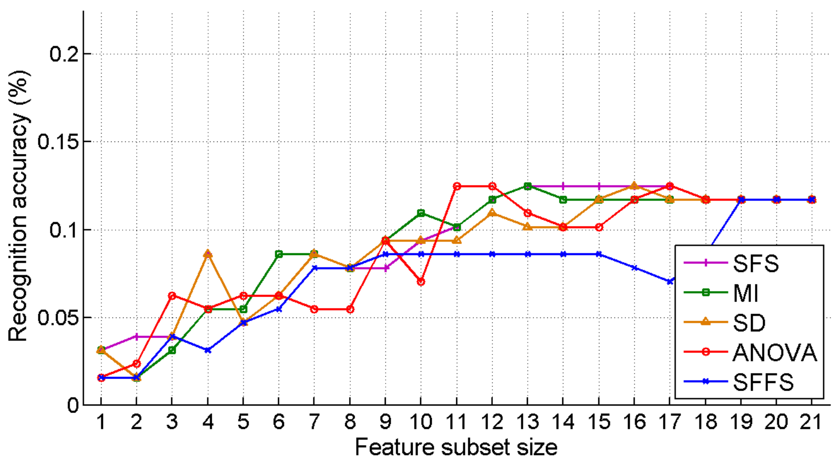


Figure 7.2: Recognition rates at **rank 1** of *ManCat-21* per method along the feature subset size increase

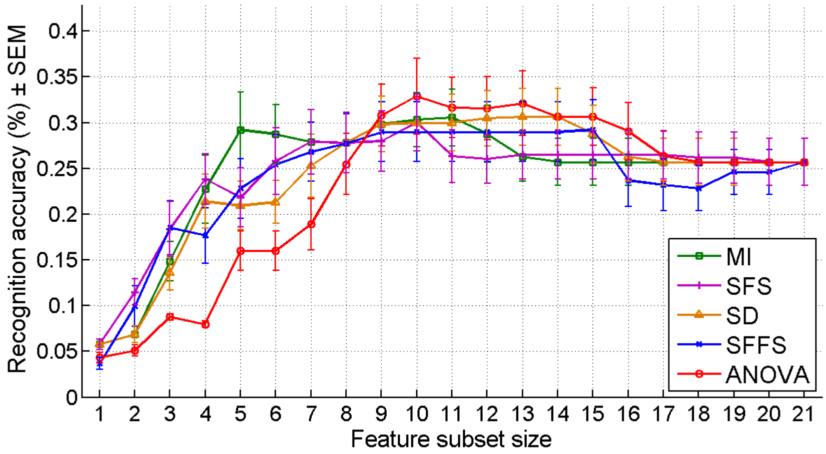


Figure 7.3: Recognition rates at **rank 10** of *ManCat-21* per method along the feature subset size increase

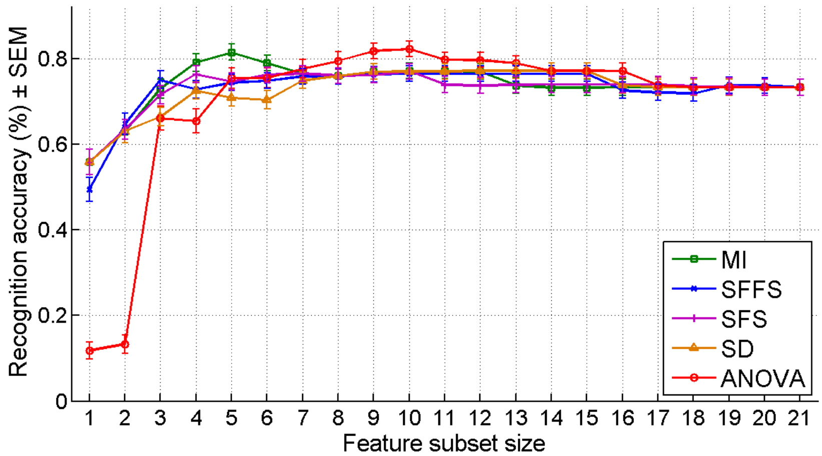


Figure 7.4: Recognition rates at **rank 128** of *ManCat-21* per method along the feature subset size increase

7.1.2 Evaluating Manual Comparative Clothing Features

Table 7.2 shows the ordering lists of manual comparative clothing traits (*ManCmp-7*) deduced using SFS and SFFS methods. In overview, Sleeve length (M8) remains here as the most powerful trait and occupies either rank 1 or 2 in the three rankings (see Section 3.3, and 3.4). M8 is also recommended as either the second or third trait in SFFS and SFS orderings respectively as in Table 7.2. The three traits M8, Head coverage (M2), and Leg length (M11) are shown as the most useful features in overall rankings and orderings. M2 is the top trait by ANOVA, M8 is the pioneer by SD and MI. M11 in the top of SFS and SFFS ordering lists. The bar chart in Fig. 7.5 represents min-max normalized ranking scores per method for *ManCmp-7* traits sorted by average score. It shows M8 as the most powerful trait followed by M2 as they receive the highest ranking scores by all means.

Table 7.2: Ordering manual comparative clothing traits using SFS and SFFS

***ManCmp-7* Soft Clothing Traits**		
Ordering	**SFS**	**SFFS**
1	M11	M11
2	M2	M8
3	M8	M14
4	M3	M7
5	M16	M2
6	M14	M3
7	M7	M16

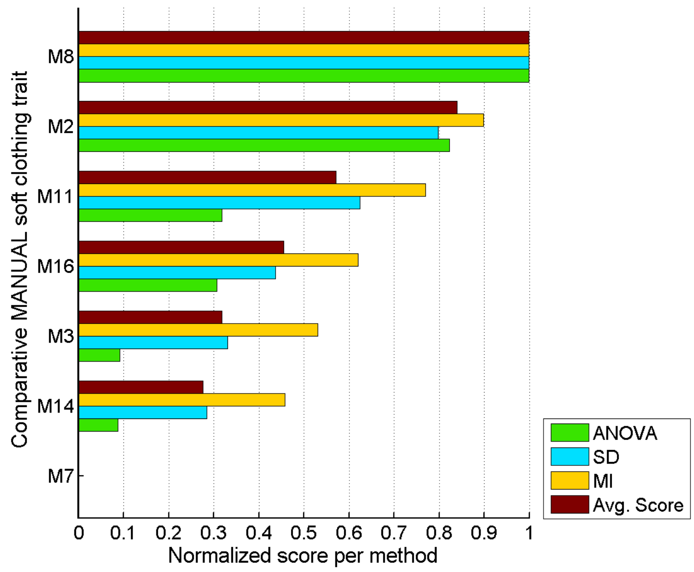


Figure 7.5: *ManCmp-7* traits ordered by average min-max normalized ranking scores to the range [0, 1]

Recognition rates using *ManCmp-7* traits are deduced at rank 1, 10, and 128 and presented in Fig. 7.6, 7.7, and 7.8 respectively, comparing the performance between the five ranking or ordering methods along all possible sizes of a feature subset (i.e. from 1 to 7). In Fig. 7.6 it is clear that all methods achieve very slight improvement in the recognition rate at rank 1 along the subset size increase from one to seven, as expected, where the best possible performance is obtained by both SD and MI using a subset consisting of only six traits and they both in addition to SFFS receive the highest average recognition overall subset sizes. In Fig. 7.7, the highest average performance and the top recognition rate using only five traits are scored by SFFS followed by ANOVA. At rank 128, as shown in Fig. 7.8, SFFS and ANOVA equally receive the highest possible performance with only six traits but ANOVA gains the best average performance overall subset sizes. In general, it can be perceived that these comparative traits contribute somewhat similarly in the recognition. Therefore having them in different rankings or orderings and using different subsets with the same size is not expected to considerably enhance the performance.

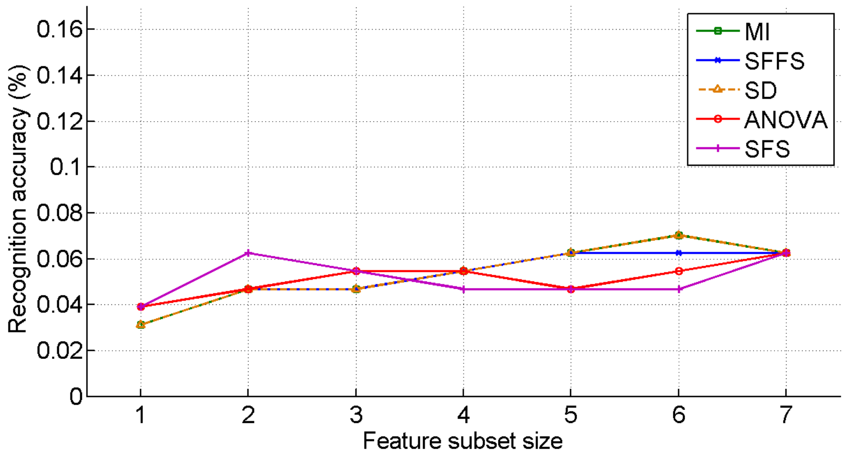


Figure 7.6: Recognition rates at **rank 1** of *ManCmp-7* per method along the feature subset size increase

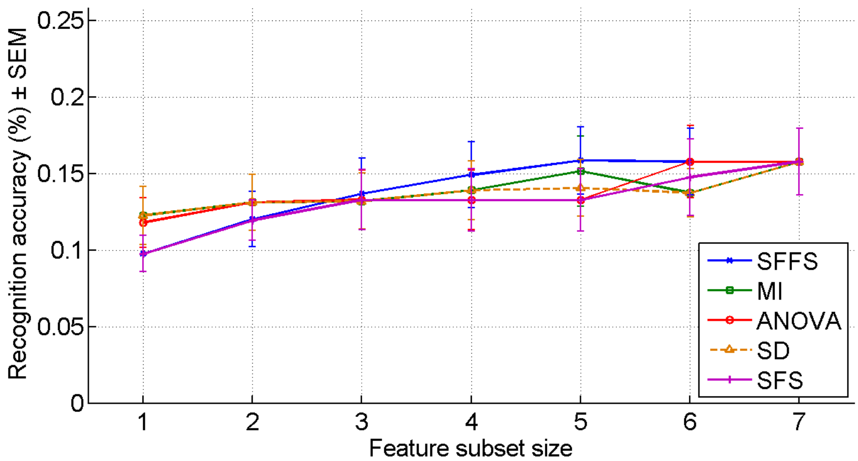


Figure 7.7: Recognition rates at **rank 10** of *ManCmp-7* per method along the feature subset size increase

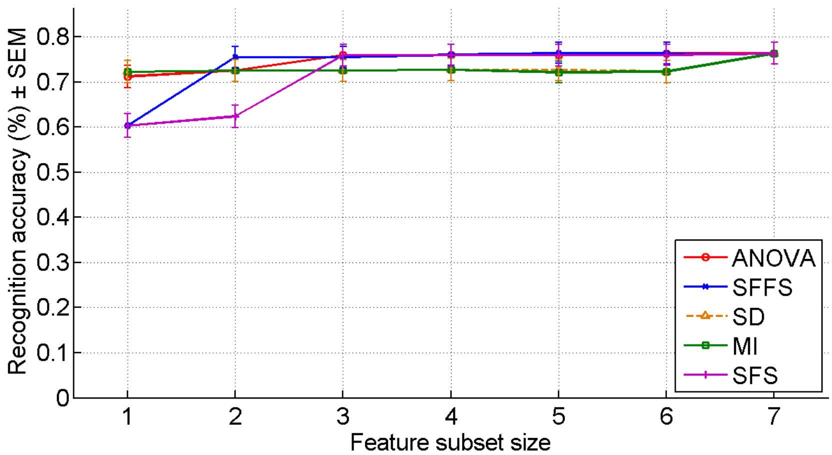


Figure 7.8: Recognition rates at **rank 128** of *ManCmp-7* per method along the feature subset size increase

7.1.3 Evaluating Automatic Categorical Clothing Features

For automatic categorical soft clothing traits (*AutoCat-17*), Table 7.3 provides different trait ordering lists obtained by SFS and SFFS methods. Besides, Fig. 7.9 shows all ranking scores of *AtuoCat-17* traits ordered by the average of min-max normalized scores per method to the range [0, 1]. The earlier reported rankings (in Section 6.5.4, and 6.5.5) show the group of traits related to brightness and contrast (A7, A13, and A3) as the top and most powerful features in overall rankings, which can be also observed at a glance in Fig. 7.9, showing these three traits with far higher scores received by all ranking methods. A7 appears as the best trait by SD and MI, while it is the third best by ANOVA. A7 is also the first suggested selection by SFFS ordering, but the third suggested selection by SFS ordering, as in Table 7.3. The traits related to colour-scheme (A8 and A14) are deemed as the next higher group in overall rankings and by the average score. Since the traits (A7, A13, A3, A8, and A14) are all semantically inferred from clothing colour information, it appears that such semantic descriptions are very informative and capable among the other clothing traits. Overall skin exposure (A1) appears better in general compared with upper (A5) and lower (A11) skin exposure.

Table 7.3: Ordering automatic categorical clothing traits using SFS and SFFS

***AutoCat-17* Soft Clothing Traits**		
Ordering	**SFS**	**SFFS**
1	A1	A7
2	A3	A14
3	A7	A8
4	A14	A9
5	A8	A11
6	A9	A13
7	A11	A10
8	A13	A1
9	A10	A15
10	A12	A12
11	A4	A3
12	A15	A4
13	A17	A16
14	A16	A5
15	A2	A2
16	A5	A17
17	A6	A6

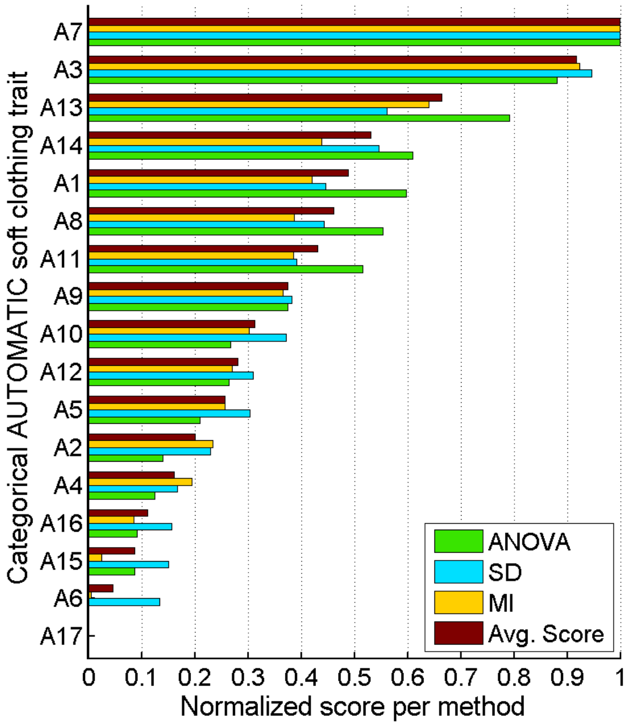


Figure 7.9: *AutoCat-17* traits ordered by average min-max normalized ranking scores to the range [0, 1]

AutoCat-17 traits are used to compare the recognition rates achieved by the five ranking or ordering methods along all possible sizes of the a feature subset (i.e. from 1 to 17) at rank 1, 10, and 128, as illustrated in Fig. 7.10, 7.11, and 7.12 respectively. Overall subset sizes from one to 17, the average performance of SFFS and then SFS consistently surpass the other methods in every respect. SD yields the highest possible score using only 15 traits at all three ranks i.e. 1, 10, and 128. In overview, the resulting curves in all cases show that the performance cumulatively increases with the subset size increase, indicating that almost all *AutoCat-17* clothing traits contribute and collaborate each other in representing a (potentially) distinctive subject identity.

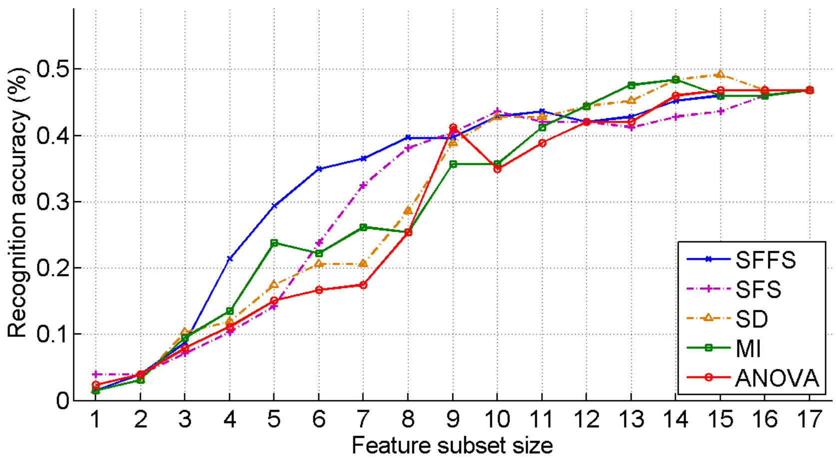


Figure 7.10: Recognition rates at **rank 1** of *AutoCat-17* per method along the feature subset size increase

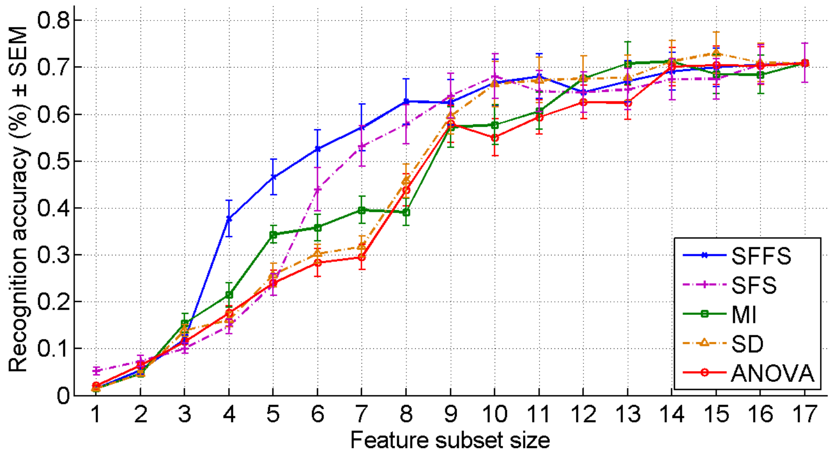


Figure 7.11: Recognition rates at **rank 10** of *AutoCat-17* per method along the feature subset size increase

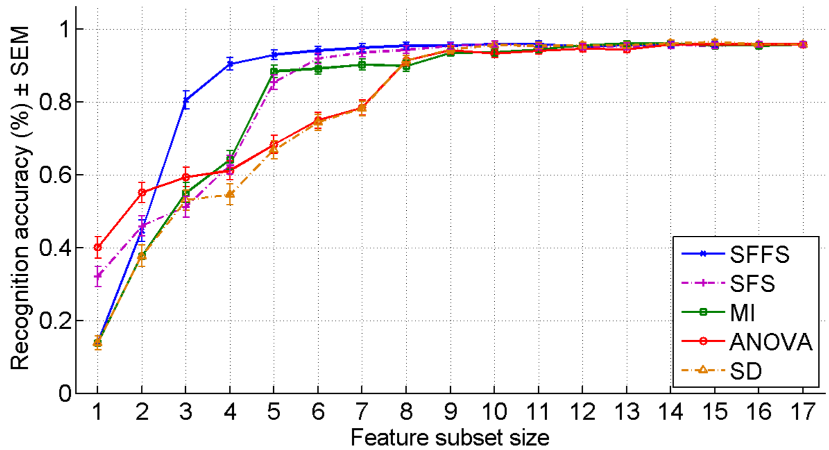


Figure 7.12: Recognition rates at **rank 128** of *AutoCat-17* per method along the feature subset size increase

7.1.4 Evaluating Automatic Comparative Clothing Features

Table 7.4 provides the ranking or ordering lists of automatic comparative soft clothing traits (*AutoCmp-10*), which are inferred using SFS and SFFS methods. Furthermore, Fig. 7.13 shows min-max normalized scores per ranking method to the range [0, 1] of all *AutoCmp-10* traits ordered by the average score (see Section 6.5.4, and 6.5.5). Upper and lower brightness (A7 and A13) and the contrast between upper and lower clothing (A3) are placed in the top of the overall rankings as in Fig. 7.13, where A7 is apparently on the top rank of ANOVA, SD, and MI rankings. Moreover, Lower skin exposure (A11) followed by upper and lower pattern (A10 and A16) are highlighted as the next powerful and useful traits based on the average scores comparison illustrated in Fig. 7.13. Although Upper dominant colour (A9) appears in the top of SFS ordering, it appears at the end of the ordering suggested by SFFS and appears very late in the tail of rankings by ANOVA, SD, and MI.

Table 7.4: Ordering automatic comparative clothing traits using SFS and SFFS

***AutoCmp-10* Soft Clothing Traits**		
Ordering	**SFS**	**SFFS**
1	A9	A13
2	A10	A11
3	A7	A3
4	A3	A10
5	A11	A7
6	A13	A16
7	A15	A15
8	A16	A1
9	A5	A5
10	A1	A9

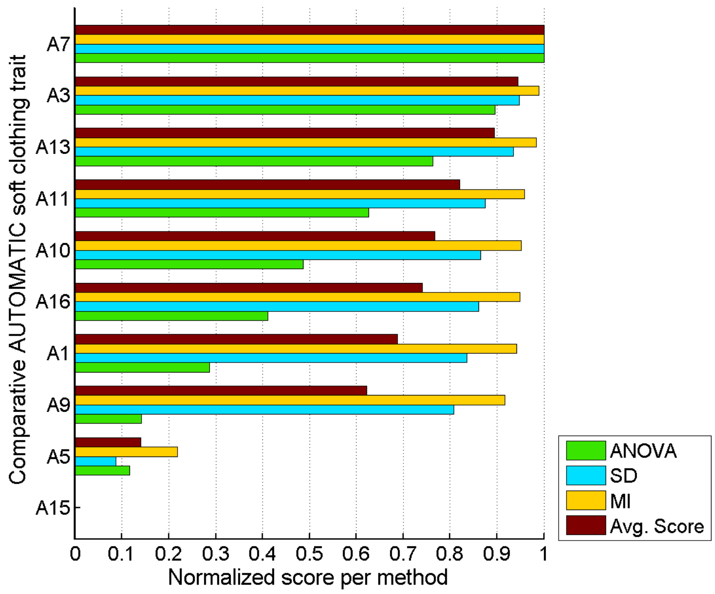


Figure 7.13: *AutoCmp-10* traits ordered by average min-max normalized ranking scores to the range [0, 1]

Fig. 7.14, 7.15, and 7.16 show recognition rates at rank 1, 10, and 128 respectively using *AutoCmp-10* clothing traits, which allow performance comparison between the five ranking or ordering methods across all possible sizes of a feature subset (i.e. from 1 to 10). At all three ranks, the ensuing curves sort the five methods from the highest to the lowest average performance as follows, SFS, ANOVA, MI, SD, and SFFS. Whilst SFS obtains similarly, at rank 1, 10, and128, the highest possible performance using a subset of only nine automatic comparative traits.

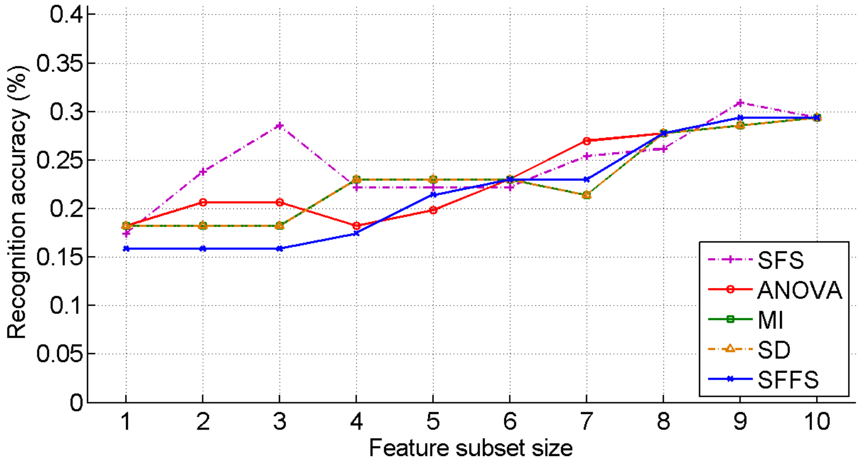


Figure 7.14: Recognition rates at **rank 1** of *AutoCmp-10* per method along the feature subset size increase

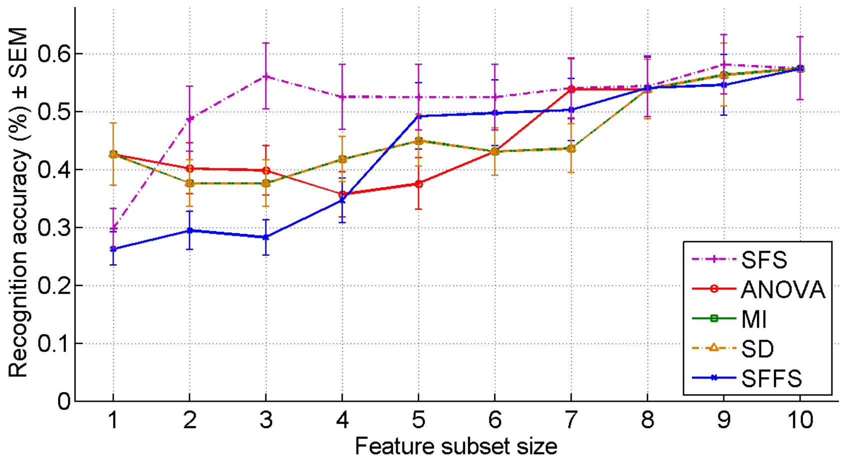


Figure 7.15: Recognition rates at **rank 10** of *AutoCmp-10* per method along the feature subset size increase

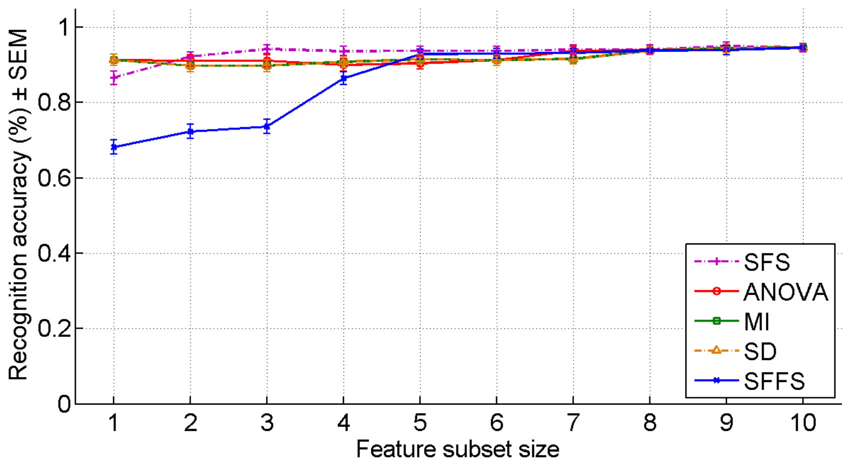


Figure 7.16: Recognition rates at **rank 128** of *AutoCmp-10* per method along the feature subset size increase

7.2 Performance Improvement via Feature Subset Selection

The main objective of applying feature selection is to determine the most informative and relevant features and to include them in a reduced feature subset achieving the highest possible identification accuracy [115]. We conduct here an empirical study aiming to utilize the feature analysis and ranking or orderings obtained in Section 7.1 to enforce effective feature subset selection via each of the five methods i.e. ANOVA, SD, MI, SFS, and SFFS. This study supplies insightful information about soft clothing traits and explores to what extent their identification performance can be enhanced, which also enables comparison between these feature selection methods.

A number of experiments are conducted on feature subset selection for enhancing biometric retrieval performance in three aspects: (i) when supplementing traditional soft biometrics by clothing; (ii) when supplementing soft body biometrics by clothing; and (iii) when using soft clothing traits alone. As such, with respect to these aspects, three experiments are achieved for each of the main soft clothing approaches (*ManCat-21*, *ManCmp-7*, *AutoCat-17*, and *AutoCmp-10*). In each experiment, the usefulness of the feature selection methods (ANOVA, SD, MI, SFS, and SFFS) is evaluated and compared with the baseline of using the full-set of clothing features alone or when added to traditional or body traits. Thus, each feature selection method is meant to choose a minimum number of soft clothing traits, which achieve the highest possible retrieval performance; namely this performance evaluation comprises (4 clothing approaches $\times$ 3 retrieval aspects $\times$ 5 feature selection methods).

Table 7.5 outlines the main manual/automatic soft clothing approaches –with the full-set of traits– used as baselines for comparison; the soft body galleries to be supplemented by clothing; and five subsets per each main clothing approach selected by the five feature selection methods. Furthermore, CMC performance scores are reported (as defined in Section 4.1.1) and consequent curves are displayed for each group of relevant clothing approaches to compare them against each other and against a baseline approach. For the sake of clarifying and magnifying the differences between CMC curves, each figure shows different limit (or maximum rank) of CMC performance, focussing only on the major part of the curves where performance changes and differences are likely notable.

Table 7.5: Automatic/manual clothing approaches and body galleries for retrieval testing and comparison

Clothing-based soft biometrics		**Description**
MANUAL	***ManCat-21***	21 *manual* categorical soft clothing traits
	ManCmp-7	7 *manual* comparative soft clothing traits
AOTOMATIC	***AutoCat-17***	17 *automatic* categorical soft clothing traits
	AutoCmp-10	10 *automatic* comparative soft clothing traits

Body-based soft biometrics	**Description**
tradSoft	4 categorical soft body biometrics (Age, Ethnicity, Sex, and Skin Colour)
softBody	17 categorical soft body biometrics including ***tradSoft*** (observable from front/side views)

Feature subset selection of soft clothing biometrics; used alone or combined with Soft body biometrics

Trait type		**Feature subset selection method**				
		ANOVA	**SD**	**MI**	**SFS**	**SFFS**
Subset of ***ManCat-21***	***& tradSoft***	***tradManCat*** **(ANOVA-6)** 6 *manual* categorical soft clothing traits with ***tradSoft***	***tradManCat*** **(SD-8)** 8 *manual* categorical soft clothing traits with ***tradSoft***	***tradManCat*** **(MI-5)** 5 *manual* categorical soft clothing traits with ***tradSoft***	***tradManCat*** **(SFS-4)** 4 *manual* categorical soft clothing traits with ***tradSoft***	***tradManCat*** **(SFFS-3)** 3 *manual* categorical soft clothing traits with ***tradSoft***
	& softBody	***softManCat*** **(ANOVA-6)** 6 *manual* categorical soft clothing traits with ***softBody***	***softManCat*** **(SD-8)** 8 *manual* categorical soft clothing traits with ***softBody***	***softManCat*** **(MI-5)** 5 *manual* categorical soft clothing traits with ***softBody***	***softManCat*** **(SFS-4)** 4 *manual* categorical soft clothing traits with ***softBody***	***softManCat*** **(SFFS-3)** 3 *manual* categorical soft clothing traits with ***softBody***
	Alone	***ManCat*** **(ANOVA-11)** Only 11 *manual* categorical soft clothing traits	***ManCat*** **(SD-12)** Only 12 *manual* categorical soft clothing traits	***ManCat*** **(MI-6)** Only 6 *manual* categorical soft clothing traits	***ManCat*** **(SFS-7)** Only 7 *manual* categorical soft clothing traits	***ManCat*** **(SFFS-10)** Only 10 *manual* categorical soft clothing traits
Subset of ***ManCmp-7***	***& tradSoft***	***tradManCmp*** **(ANOVA-3)** 3 *manual* comparative soft clothing traits with ***tradSoft***	***tradManCmp*** **(SD-4)** 4 *manual* comparative soft clothing traits with ***tradSoft***	***tradManCmp*** **(MI-4)** 4 *manual* comparative soft clothing traits with ***tradSoft***	***tradManCmp*** **(SFS-6)** 6 *manual* comparative soft clothing traits with ***tradSoft***	***tradManCmp*** **(SFFS-5)** 5 *manual* comparative soft clothing traits with ***tradSoft***
	& softBody	***softManCmp*****(ANOVA-3)** 3 *manual* comparative soft clothing traits with ***softBody***	***softManCmp*****(SD-4)** 4 *manual* comparative soft clothing traits with ***softBody***	***softManCmp*****(MI-4)** 4 *manual* comparative soft clothing traits with ***softBody***	***softManCmp*****(SFS-6)** 6 *manual* comparative soft clothing traits with ***softBody***	***softManCmp*****(SFFS-5)** 5 *manual* comparative soft clothing traits with ***softBody***
	Alone	***ManCmp*** **(ANOVA-3)** Only 3 *manual* comparative soft clothing traits	***ManCmp*** **(SD-4)** Only 4 *manual* comparative soft clothing traits	***ManCmp*** **(MI-4)** Only 4 *manual* comparative soft clothing traits	***ManCmp*** **(SFS-6)** Only 6 *manual* comparative soft clothing traits	***ManCmp*** **(SFFS-5)** Only 5 *manual* comparative soft clothing traits
Subset of ***AutoCat-17***	***& tradSoft***	***tradAutoCat*** **(ANOVA-14)** 14 *automatic* categorical clothing traits with ***tradSoft***	***tradAutoCat*** **(SD-16)** 16 *automatic* categorical clothing traits with ***tradSoft***	***tradAutoCat*** **(MI-16)** 16 *automatic* categorical clothing traits with ***tradSoft***	***tradAutoCat*** **(SFS-14)** 14 *automatic* categorical clothing traits with ***tradSoft***	***tradAutoCat*** **(SFFS-14)** 14 *automatic* categorical clothing traits with ***tradSoft***
	& softBody	***softAutoCat*** **(ANOVA-14)** 14 *automatic* categorical clothing traits with ***softBody***	***softAutoCat*** **(SD-16)** 16 *automatic* categorical clothing traits with ***softBody***	***softAutoCat*** **(MI-16)** 16 *automatic* categorical clothing traits with ***softBody***	***softAutoCat*** **(SFS-14)** 14 *automatic* categorical clothing traits with ***softBody***	***softAutoCat*** **(SFFS-14)** 14 *automatic* categorical clothing traits with ***softBody***
	Alone	***AutoCat*** **(ANOVA-14)** Only 14 *automatic* categorical clothing traits	***AutoCat*** **(SD-16)** Only 16 *automatic* categorical clothing traits	***AutoCat*** **(MI-16)** Only 16 *automatic* categorical clothing traits	***AutoCat*** **(SFS-14)** Only 14 *automatic* categorical clothing traits	***AutoCat*** **(SFFS-14)** Only 14 *automatic* categorical clothing traits
Subset of ***AutoCmp-10***	***& tradSoft***	***tradAutoCmp*** **(ANOVA-5)** 5 *automatic* comparative clothing traits with ***tradSoft***	***tradAutoCmp*** **(SD-6)** 6 *automatic* comparative clothing traits with ***tradSoft***	***tradAutoCmp*** **(MI-6)** 6 *automatic* comparative clothing traits with ***tradSoft***	***tradAutoCmp*** **(SFS-6)** 6 *automatic* comparative clothing traits with ***tradSoft***	***tradAutoCmp*** **(SFFS-6)** 6 *automatic* comparative clothing traits with ***tradSoft***
	& softBody	***softAutoCmp*****(ANOVA-5)** 5 *automatic* comparative clothing traits with ***softBody***	***softAutoCmp*****(SD-6)** 6 *automatic* comparative clothing traits with ***softBody***	***softAutoCmp*****(MI-6)** 6 *automatic* comparative clothing traits with ***softBody***	***softAutoCmp*****(SFS-6)** 6 *automatic* comparative clothing traits with ***softBody***	***softAutoCmp*****(SFFS-6)** 6 *automatic* comparative clothing traits with ***softBody***
	Alone	***AutoCmp*** **(ANOVA-5)** Only 5 *automatic* comparative clothing traits	***AutoCmp*** **(SD-6)** Only 6 *automatic* comparative clothing traits	***AutoCmp*** **(MI-6)** Only 6 *automatic* comparative clothing traits	***AutoCmp*** **(SFS-6)** Only 6 *automatic* comparative clothing traits	***AutoCmp*** **(SFFS-6)** Only 6 *automatic* comparative clothing traits

7.2.1 Feature Selection for Manual Categorical Clothing Approach

Table 7.6 reports CMC scores of *ManCat-21* traits and compares the capabilities of different subsets selected by the adopted feature selection methods i.e. ANOVA, SD, MI, SFS, and SFFS. Accordingly, the CMC performance is grouped and compared per retrieval aspects, where the performance is shown when clothing subsets are used with traditional traits in Fig. 7.17 and with body traits in Fig. 7.18, or when used alone in Fig. 7.19. In general, ANOVA achieves the best overall performance in all retrieval aspects. MI is very competitive to ANOVA, especially when used with *tradSoft* as in Fig. 7.17 or when used alone as in Fig. 7.19. Moreover, it offers the second best overall performance in terms of supplementing *tradSoft* and *softBody*, whereas a subset containing only six clothing traits selected by SFS attains the second best performance when used alone. It can be observed in all figures that all used feature selection methods can significantly enhance the baseline performance (shown in black) using a reduced feature subset of manual categorical clothing traits.

Table 7.6: CMC scores of retrieval using the feature selection methods on *ManCat-21* traits

Retrieval aspect (Trait type)	**Approach**	**Top rank**	**AVG sum match scores up to rank**		**100% accuracy achieved at rank**	**Overall rank**
		=1	***=10***	***=128***		
Clothing & ***tradSoft***	*tradSoft*	0.187	0.342	0.872	72	6
	tradManCat (ANOVA-6)	**0.318**	0.595	0.940	**57**	**1**
	tradManCat (SD-8)	0.266	0.556	0.925	62	3
	tradManCat (MI-6)	0.313	**0.630**	**0.942**	64	2
	tradManCat (SFS-4)	0.281	0.563	0.931	67	4
	tradManCat (SFFS-3)	0.254	0.554	0.929	90	5
Clothing & ***softBody***	*softBody*	0.858	0.952	0.9949	30	6
	softManCat (ANOVA-6)	**0.930**	**0.978**	**0.9973**	27	**1**
	softManCat (SD-8)	0.867	0.962	0.9963	**23**	4
	softManCat (MI-6)	0.906	0.971	0.9969	25	2
	softManCat (SFS-4)	0.914	0.976	0.9971	27	3
	softManCat (SFFS-3)	0.914	0.971	0.9966	29	5
Clothing ***Alone***	*ManCat-21*	0.125	0.238	0.762	116	6
	ManCat (ANOVA-11)	0.102	**0.347**	**0.842**	102	**1**
	ManCat (SD-12)	0.133	0.306	0.818	124	5
	ManCat (MI-6)	0.094	0.304	0.833	106	3
	ManCat (SFS-7)	0.102	0.306	0.816	**100**	2
	ManCat (SFFS-10)	**0.141**	0.323	0.810	111	4

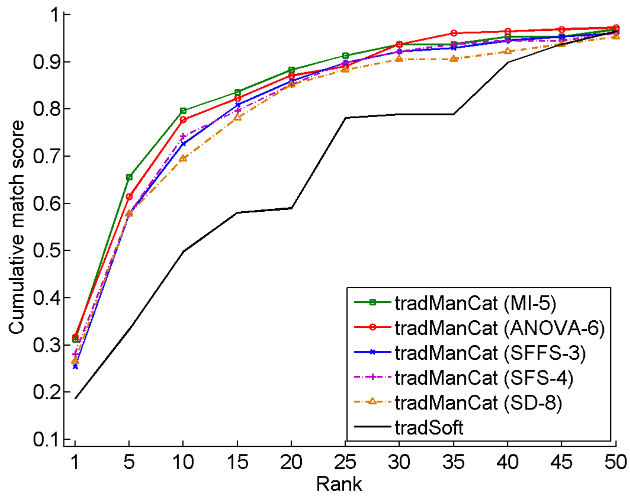


Figure 7.17: Performance comparison between the feature selection methods on *ManCat-21* added to *tradSoft*

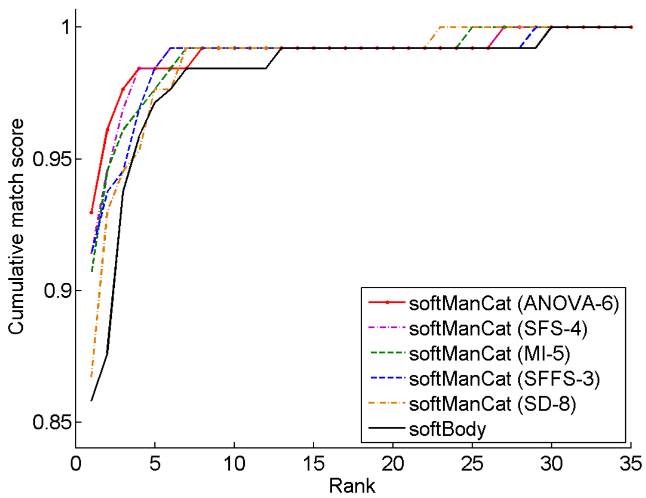


Figure 7.18: Performance comparison between the feature selection methods on *ManCat-21* added to *softBody*

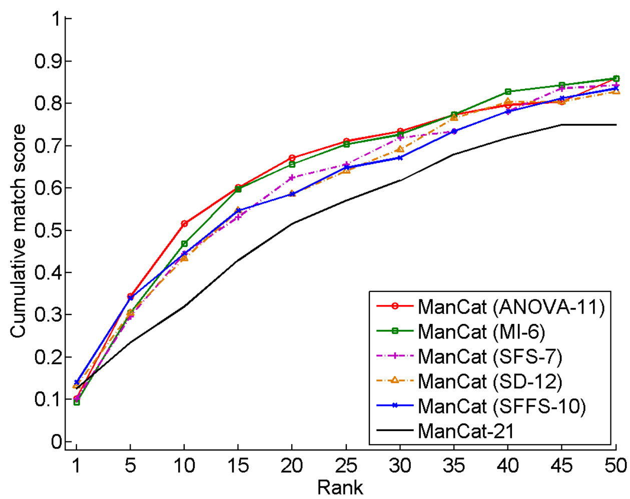


Figure 7.19: Performance comparison between the feature selection methods on *ManCat-21* when used *alone*

7.2.2 Feature Selection for Manual Comparative Clothing Approach

For retrieval by *ManCmp-7* traits and for comparing their performance using the feature selection methods, Table 7.7 presents associated CMC scores for all retrieval aspects. Besides, the corresponding CMC curves are shown in Fig. 7.20, 7.21, and 7.22. SFFS proposes a subset of only five traits and surpasses all methods, in terms of retrieval by clothing and *tradSoft*, as indicated in Table 7.7 and also reflected by Fig. 7.20. Furthermore, SFFS subset is also the best in performance when used alone. In contrast it appears the lowest in performance when used with *softBody*, while MI and SD are equally using the same feature subset of only four traits and scoring the highest retrieval accuracy, as in Fig. 7.21. In the use of clothing traits alone, the highest possible retrieval performance by SD and MI subsets does not surpass the baseline performance that uses the full set of *ManCmp-7* traits, as can be seen in the table scores and in Fig. 7.22.

Table 7.7: CMC scores of retrieval using the feature selection methods on *ManCmp-7* traits

Retrieval aspect (Trait type)	**Approach**	**Top rank**	**AVG sum match scores up to rank**		**100% accuracy achieved at rank**	**Overall rank**
		=1	***=10***	***=128***		
Clothing & ***tradSoft***	*tradSoft*	0.187	0.342	0.872	72	6
	tradManCmp (ANOVA-3)	**0.281**	0.584	0.941	64	2
	tradManCmp (SD-4)	0.273	0.549	0.929	65	4
	tradManCmp (MI-4)	0.273	0.549	0.929	65	4
	tradManCmp (SFS-6)	0.273	0.583	0.941	64	3
	tradManCmp (SFFS-5)	**0.281**	**0.588**	**0.944**	**63**	**1**
Clothing & ***softBody***	*softBody*	0.858	0.952	0.9949	**30**	6
	softManCmp (ANOVA-3)	0.914	0.971	0.9965	**30**	3
	softManCmp (SD-4)	**0.922**	**0.973**	**0.9966**	**30**	**1**
	softManCmp (MI-4)	**0.922**	**0.973**	**0.9966**	**30**	**1**
	softManCmp (SFS-6)	0.914	0.972	**0.9966**	**30**	3
	softManCmp (SFFS-5)	0.906	0.970	0.9964	**30**	5
Clothing ***Alone***	*ManCmp-7*	**0.059**	0.160	0.769	**105**	2
	ManCmp (ANOVA-3)	0.051	0.134	0.765	111	4
	ManCmp (SD-4)	0.051	0.139	0.732	119	5
	ManCmp (MI-4)	0.051	0.139	0.732	119	5
	ManCmp (SFS-6)	0.043	0.151	0.765	112	3
	ManCmp (SFFS-5)	**0.059**	**0.161**	**0.770**	**105**	**1**

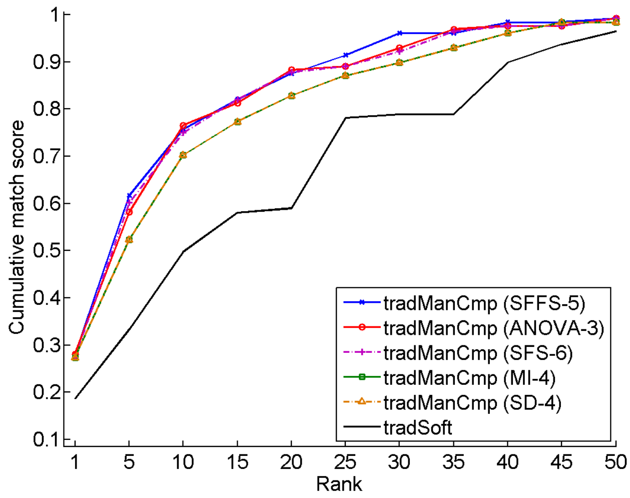


Figure 7.20: Performance comparison between the feature selection methods on *ManCmp-7* added to *tradSoft*

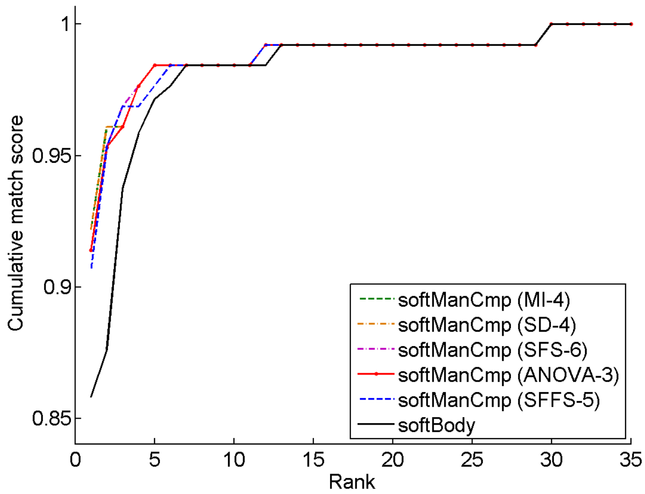


Figure 7.21: Performance comparison between the feature selection methods on *ManCmp-7* added to *softBody*

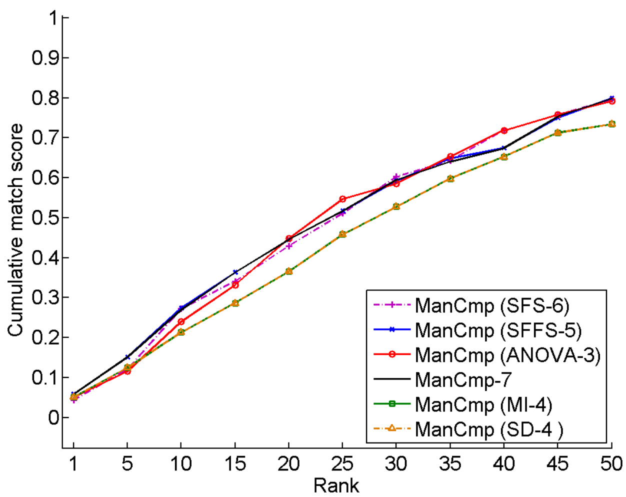


Figure 7.22: Performance comparison between the feature selection methods on *ManCmp-7* when use *alone*

7.2.3 Feature Selection for Automatic Categorical Clothing Approach

Table 7.8 provides CMC scores for each retrieval aspect and compares the capabilities of different *AutoCat-17* subsets, which are nominated by the feature selection methods. Whilst Fig. 7.23, 7.24, and 7.25 illustrate and compare the performance of these subsets when used with *tradSoft*, *softBody*, and alone respectively. In overview, all feature selection methods achieve very similar performance in every respect and use in their selected subsets a large number of traits (ranging from 14 to 16) close to the full size of *AutoCat-17*. This may indicate that almost all traits are useful for retrieval and they significantly contribute and cooperate to construct a distinctive biometric identity. As in Table 7.8 and Fig. 7.23, all methods considerably exceed the performance of the baseline with very slight differences between them, where SFS receives the top scores. ANOVA performs the best in case of using *AutoCat-17* subsets with *softBody* and the remaining methods still performs much better than the baseline shown in black in Fig. 7.24. Even when these subsets are used alone, they all but MI enhance the performance of the baseline i.e. *AutoCat-17* embracing all traits, where SFFS is the best in every respect.

Table 7.8: CMC scores of retrieval using the feature selection methods on *AutoCat-17* traits

Retrieval aspect (Trait type)	**Approach**	**Top rank**	**AVG sum match scores up to rank**		**100% accuracy achieved at rank**	**Overall rank**
		=1	***=10***	***=128***		
Clothing & ***tradSoft***	*tradSoft*	0.187	0.342	0.8719	72	6
	tradAutoCat (ANOVA-14)	**0.729**	0.898	0.9910	**21**	3
	tradAutoCat (SD-16)	0.721	0.890	0.9896	**21**	4
	tradAutoCat (MI-16)	0.721	0.889	0.9895	**21**	5
	tradAutoCat (SFS-14)	0.741	**0.910**	**0.9921**	**21**	**1**
	tradAutoCat (SFFS-14)	**0.729**	0.904	0.9915	**21**	2
Clothing & ***softBody***	*softBody*	0.858	0.952	0.9949	30	6
	softAutoCat (ANOVA-14)	0.992	**0.999**	**0.9999**	**2**	**1**
	softAutoCat (SD-4)	**0.977**	0.997	0.9998	**2**	4
	softAutoCat (MI-4)	**0.977**	0.997	0.9998	**2**	4
	softAutoCat (SFS-14)	0.992	0.998	**0.9999**	3	2
	softAutoCat (SFFS-14)	0.992	0.998	**0.9999**	3	2
Clothing ***Alone***	*AutoCat-17*	0.4943	0.724	0.9642	44	3
	AutoCat (ANOVA-14)	0.4950	0.740	0.9681	44	2
	AutoCat (SD-16)	0.4943	0.725	0.9642	45	5
	AutoCat (MI-16)	0.4940	0.718	0.9636	45	6
	AutoCat (SFS-14)	0.4908	0.733	0.9695	46	4
	AutoCat (SFFS-14)	**0.4976**	**0.743**	**0.9700**	**39**	**1**

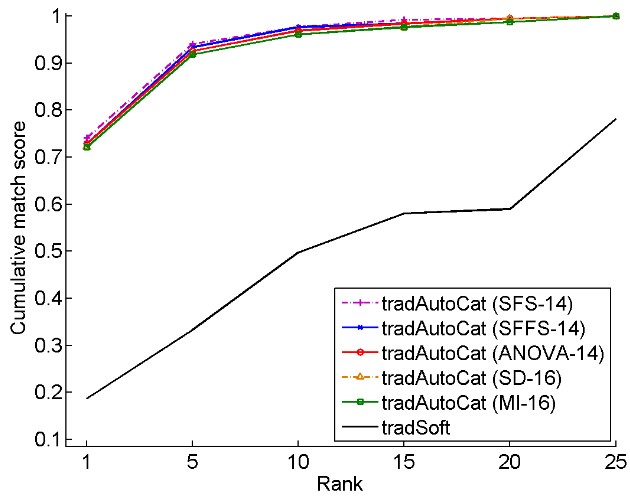


Figure 7.23: Performance comparison between the feature selection methods on *AutoCat-17* added to *tradSoft*

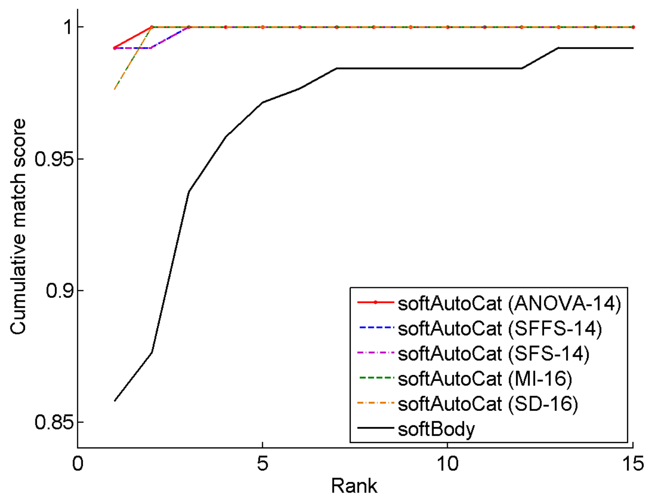


Figure 7.24: Performance comparison between the feature selection methods on *AutoCat-17* added to *softBody*

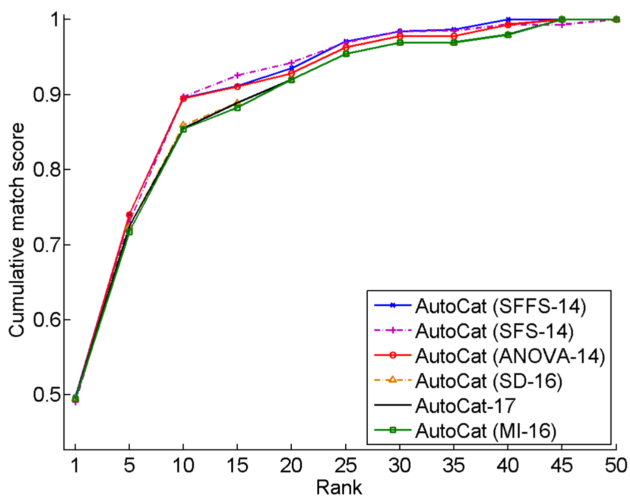


Figure 7.25: Performance comparison between the feature selection methods on *AutoCat-17* when used *alone*

7.2.4 Feature Selection for Automatic Comparative Clothing Approach

The feature selection methods are applied on *AutoCmp-10* and compared in terms of their best possible performance in all retrieval aspects as given in Table 7.9 and demonstrated for use with *tradSoft*, *softBody*, and alone in Fig. 7.26, 7.27, and 7.28 respectively. A subset containing only six traits selected by SFFS is reported as the highest performance method when used with *tradSoft* or alone, as demonstrated in Fig. 7.26 and 7.28. SD and MI perform exactly the same with a subset of six traits, since they have the same ordering for *AutoCmp-10* traits, and they attain the highest scores and the best overall performance when supplementing *softBody* traits as shown in Fig. 7.27. In both Fig. 7.26 and 7.28 along with the corresponding scores in Table 7.9, it can be observed that all methods have very similar trends in retrieval performance, where they greatly outperform *tradSoft* and slightly improve *AutoCmp-10* when used alone.

Table 7.9: CMC scores of retrieval using the feature selection methods on *AutoCmp-10* traits

Retrieval aspect (Trait type)	**Approach**	**Top rank**	**AVG sum match scores up to rank**		**100% accuracy achieved at rank**	**Overall rank**
		=1	*=10*	*=128*		
Clothing & ***tradSoft***	*tradSoft*	0.187	0.342	0.8719	72	6
	tradAutoCmp (ANOVA-5)	**0.665**	0.891	0.9872	62	2
	tradAutoCmp (SD-6)	0.642	0.889	0.9875	**59**	3
	tradAutoCmp (MI-6)	0.642	0.889	0.9875	**59**	3
	tradAutoCmp (SFS-6)	0.621	0.890	0.9876	**59**	5
	tradAutoCmp (SFFS-6)	0.652	**0.894**	**0.9879**	**59**	**1**
Clothing & ***softBody***	*softBody*	0.858	0.9519	0.9949	30	6
	softAutoCmp (ANOVA-5)	0.961	0.9844	0.9981	22	5
	softAutoCmp (SD-6)	**0.977**	**0.9875**	**0.9984**	22	**1**
	softAutoCmp (MI-6)	**0.977**	**0.9875**	**0.9984**	22	**1**
	softAutoCmp (SFS-6)	0.953	0.9836	0.9982	**20**	4
	softAutoCmp (SFFS-6)	0.961	0.9844	0.9982	**20**	3
Clothing ***Alone***	*AutoCmp-10*	0.296	0.588	0.949	69	6
	AutoCmp (ANOVA-5)	**0.347**	0.612	0.954	54	3
	AutoCmp (SD-6)	0.291	0.600	0.952	62	4
	AutoCmp (MI-6)	0.291	0.600	0.952	62	4
	AutoCmp (SFS-6)	0.333	0.619	**0.956**	**53**	2
	AutoCmp (SFFS-6)	0.343	**0.623**	**0.956**	**53**	**1**

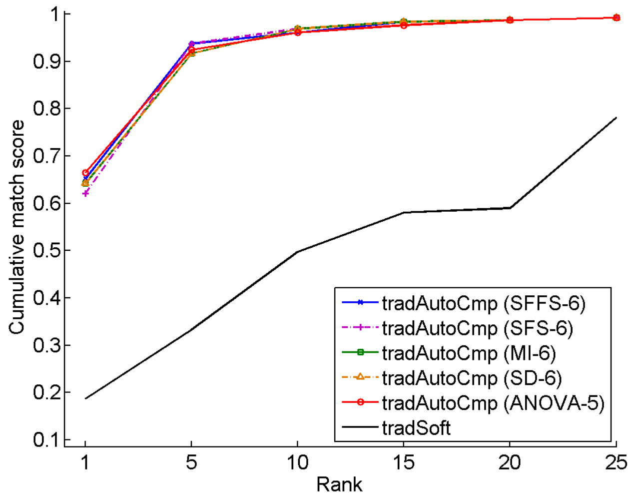


Figure 7.26: Performance comparison between the feature selection methods on *AutoCmp*-10 added to *tradSoft*

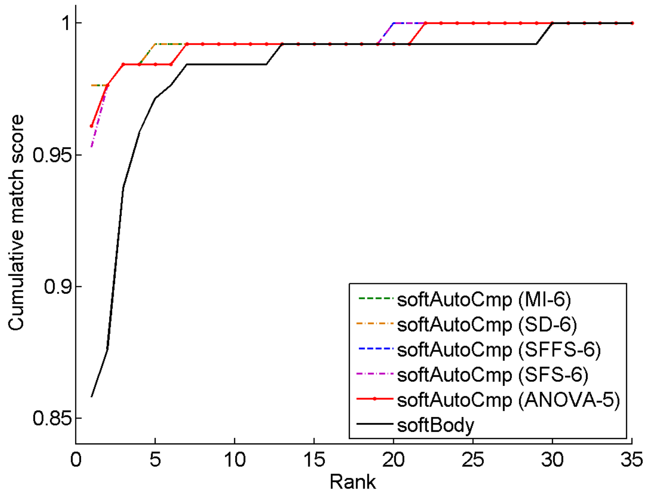


Figure 7.27: Performance comparison between the feature selection methods on *AutoCmp*-10 added to *softBody*

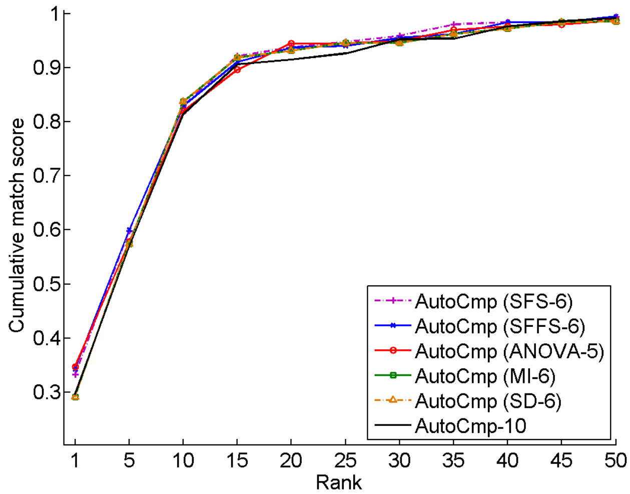


Figure 7.28: Performance comparison between the feature selection methods on *AutoCmp*-10 when used *alone*

Eventually, since we have already ranked the competitive feature selection methods in the preceding empirical study, it is worth to deduce an (overall) final ranking for these methods, highlighting the most successful method in selecting the most effective traits from a variety of soft clothing biometric forms and approaches. Table 7.10, summarises the ranking results for the five feature selection methods over the earlier experiments and accomplishes with a final rank for each method. Hence, it is obvious that, ANOVA as a feature subset selection method surpasses the other methods by mostly selecting a smaller number of highly effective and very discriminative soft clothing traits. SFFS is the next best method and appears to be more suited to our context than SFS.

Table 7.10: Ranking feature selection methods based on performance across experiments

Method	***ManCat-21***			***ManCmp-7***			***AutoCat-17***			***AutoCmp-10***			**Final rank**
	& tradSoft	***& softBody***	***Alone***	***& tradSoft***	***& softBody***	***Alone***	***& tradSoft***	***& softBody***	***Alone***	***& tradSoft***	***& softBody***	***Alone***	
ANOVA	**1**	**1**	**1**	2	3	4	3	**1**	2	2	5	3	**1**
SD	3	4	5	4	**1**	5	4	4	5	3	**1**	4	5
MI	2	2	3	4	**1**	5	5	4	6	3	**1**	4	4
SFS	4	3	2	3	3	3	**1**	2	4	5	4	2	3
SFFS	5	5	4	**1**	5	**1**	2	2	**1**	**1**	3	**1**	2

7.3 Conclusions

This chapter provides extended analysis and empirical study on the capabilities of soft biometrics using clothing attributes derived manually or automatically for identification and retrieval purposes. All the main forms of soft clothing traits are listed based on retrieval performance using three scoring/ranking methods ANOVA, SD, and MI in addition to two feature selection/ordering methods SFS, and SFFS. Each method is found to be differently treating and prioritising clothing traits resulting in mostly different trait ordering suggested by each method for same list of traits. This appears indicating some potential trade-offs associated with the capabilities of soft clothing attributes. Namely if some traits are indicated as powerful traits in terms of a particular method or in the use of a particular retrieval scenario (or aspect), they may not be considered as powerful in terms of another method or when used in different retrieval scenario. For instance, a single trait may be found in the top of the ranking list by a certain method (e.g. ANOVA) but at the meantime, it is not found powerful by another method (e.g. MI), and this trait may be very viable when supplementing traditional soft biometrics (*tradSoft*) but not considered as viable when supplementing soft body biometrics (*softBody*) or when used alone.

It is shown that the retrieval performance of soft clothing biometrics can indeed be improved via different feature selection methods. Based on the experimental results, ANOVA as a feature subset selection method achieves the highest overall performance throughout the conducted retrieval experiments.

As a major conclusion of the conducted empirical study on the clothing feature space and as it may be observed and reported by earlier studies [92, 94, 115, 116], there is no definite feature selection method or feature subset size to be always the optimal subset selection or to guarantee the best performance in all cases. This is due to the fact that features may exchange their ranks or orders (or degrees of significance) because of some influencing factors, which can positively or negatively affect the potency and usefulness of these features. With regard to biometric traits, common influencing factors on feature significance ordering include distinctiveness, collectability, universality, viewpoint invariance, permanency and resistance to circumvention [26]. Changing the database (sample size, imaging viewpoint, quality, environment etc.) is perhaps a typical example where such factors may take place in changing the feature significance ordering and the data distributions of soft biometric traits, which may vary from a database to another. This in turn may affect the efficacy and usefulness of some those biometric traits. Our study may be a good empirical and practical methodology to decide or recommend the most likely effective and useful soft biometric traits.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

This thesis introduces the concept of ***soft clothing biometrics***, the first approach for identifying people by their clothing as a major cue. Two different groups of soft clothing biometrics are analysed and 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.

In this thesis, Chapter 1 introduces the research with motivations, context and contributions then provides a brief overview about human identification, soft biometrics, clothing and their semantic attributes across a number of related work. Chapter 2 describes proposed semantic clothing attributes and descriptive labels for annotation, human based clothing label data acquisition, and resulting soft clothing biometrics in categorical and comparative forms and the methodology used to derive these soft biometrics. Chapter 3 briefly builds a common understanding of the significance, differences, correlations, and efficacy of soft clothing traits in identification, through a series of statistical analysis methods. Chapter 4 experimentally investigates the power and evaluates the performance of soft clothing biometrics in human identification in case of addition to other soft biometrics or when used alone. Chapter 5 presents the use of soft clothing biometrics for enhanced subject retrieval and how they can be generalized for unseen data. It also studies the utility of soft clothing traits in more challenging biometric retrieval in realistic scenarios and examines their capability in achieving viewpoint invariant subject retrieval. Chapter 6 provides a brief description of proposed automatic soft clothing attributes and explores the methodology used for vision-based automatic extraction and annotation. It then presents the ensuing automatic soft clothing biometrics and the detailed analysis of these traits. Eventually it demonstrates a number of experiments on viewpoint invariant retrieval, investigating the power of automatic soft clothing traits derived via computer-vision and allowing performance comparison against those manual soft clothing traits inferred by humans. Chapter 7 studies the clothing feature space via extended analysis and empirical investigation of the capabilities of soft biometrics using clothing attributes in human identification and retrieval, leading to an insightful guide for feature subset selection and enhanced performance.

Clothing characteristics can be utilized to convey effective descriptions and useful combinations of soft biometrics. In some cases, human clothing descriptions might be the only observable attributes and a beneficial clue for identity.

In our study, we find that relative attributes may carry the most significant and effective information to describe clothing in either categorical or comparative form. Comparative traits are capable to represent the very small differences between subjects and to adjust the descriptions of a subject with respect to all other subjects in the database. A strong correlation between two attributes allows for predicting a missing attribute of the two most likely from the other. Further investigation of correlations between automatic and manual attributes may lead to automatic inference to those manual attributes rather than human-based annotations if not available.

The obtained performance results using semantic clothing attributes indicate a motivation for exploiting derived clothing traits, in fusion or even in isolation, to enrich human identification and re-identification. The proposed techniques enforce an identification task by using clothing attributes, in supplement to traditional or body soft traits, and as the only biometrics. We enable the implicit use of the rich correlations of clothing attributes. We show that, as a proof-of-concept, we can recognize people using soft clothing traits and we need further to learn a soft attribute-centric for more powerful identification and retrieval in the future.

Our study explores also the viability of using soft clothing attributes to achieve enhanced subject retrieval. The results of this exploration using clothing traits highlights a potentially valuable addition to the field of soft biometrics. This can lead to new and useful enhanced biometric applications and systems, using soft clothing biometrics for various purposes including subject search, retrieval, identification, and re-identification. Our analysis of soft clothing traits indicates that such clothing characteristics can be associated in biometric signatures and achieve successful subject retrieval. We show further how soft clothing biometrics can be used in more challenging retrieval scenarios and how such soft clothing traits can be immune change from front to side viewpoint.

We show that it is possible to use manually collected labels and automatically derived clothing attributes for human recognition purposes, signifying a precursor to automated search by clothing for forensic use by either or both techniques. Our proposed clothing-based approaches enable 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, 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.

In this research, manual and automatic soft clothing traits are used in a manner to bridge the semantic gap between machine and human, whereas automatic soft clothing traits are intended further to bridge another semantic gap between vision-based low-level features and high-level clothing annotations. Since this initial study aims to convey an important message that clothing information can be differentiated as soft biometrics, as part of a program aimed to achieve the capability to automatically search video and images in the wild for identification purposes.

8.2 Future Work

Proceeding from this thesis, the focus will be basically on the investigation of to what extent can clothing attributes be deployed in improved identification and retrieval, and in what sense can we refine soft clothing biometrics to suit new possible applications. A number of promising research directions arise for future work. Here we highlight some of what appear to be of most relevant interest as follows.

- **Gender Discrimination by Clothing**

Human gait, body and face information have been widely used for gender identification in most of existing research [117]. Soft facial metrology has been examined and exploited in gender classification [49], while it becomes more challenging problem in some situations such as facial pose [118]. Since humans are usually able to infer the gender of a person using some body features such as body shape and hair style, they also use additional cues like clothes types or styles and accessories [119]. Clothing information as cues are less affected by pose and illumination than faces, and appear ideal for gender detection via attribute learning [73]. Although, these cues are simply usable and effortlessly observable, the recognition task may become more difficult and confusing since people of the same gender can vary greatly in their choices of clothing styles [120]. Another challenge can exist in the common clothing styles worn by both males and females which may open a room for ambiguity or confused similarity.

In this research our study of the correlations between the soft clothing attributes can be used to pave the way for gender discrimination. As we could be able to infer the label (“Male” or “Female”) for the gender attribute from other available attributes that highly correlated with this attribute. Moreover, we can support a (confusion) decision and utilize information derived from correlations to increase or decrease the correctness probability of that decision. Gender as a soft trait has been already used as an input given to a classifier to identify the person. However, assigning gender labels as a desired output received by using soft clothing biometrics is yet to be performed. This spurs the interest to explore the perspicacity of soft clothing traits in gender prediction and in what extent the accuracy and performance can be optimized.

- **Soft Clothing Attributes and Their Applications at a Distance**

Clothing is likely to offer some soft attributes which can be observable and describable at a distance and they are expected to effectively participate in representing the basic individual's appearance and structuring the soft biometric signature. Such clothing attributes may enable desirable optimization in several applications, for example, allowing to locate quickly and correctly a person-of-interest in real time depending on a given witness descriptions, another application can be supplementing and supporting enhanced speed video-based forensic investigations [42]. As a matter of fact, the usability of soft biometrics at a distance and their robustness against degraded resolution [40] seem to be more likely in the domain of soft clothing traits, and seem to provide a motivation for prospective high performance and reliable resistance against the inseparable challenges of recognition at a distance.

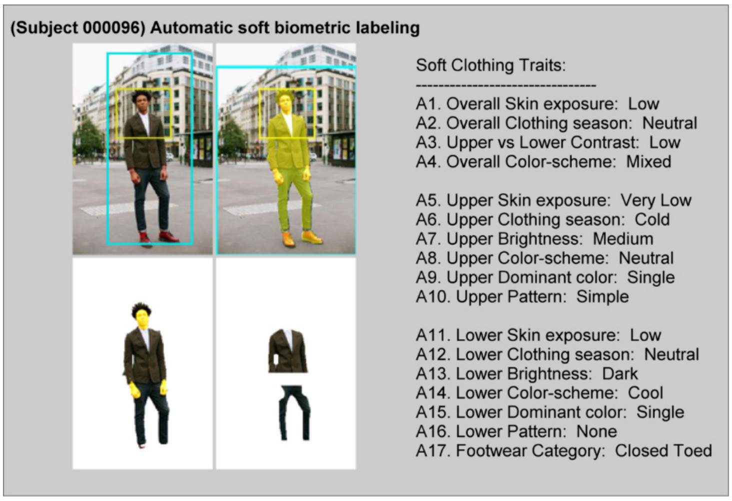
- **Verbal People Search and Retrieval from Surveillance Data**

Instead of only using verbal descriptions to match and recognize a number of particular subjects whom information is enrolled in the database, the next milestone is to learn for more powerful approaches capable to search online/offline data images and videos to retrieve all best matching persons. Hence, similar approaches may result in a wide variety of useful applications enable offline/online retrieval from imaging and footage using solely verbal multi-attribute descriptions of a person-of-interest. For instance, using verbal descriptions obtained from witnesses can be used to form a query to search and locate or retrieve a wanted suspect from available surveillance data. We aim to use as much information as we can in surveillance data, where clothes can play an important role due to their highly observable soft traits.

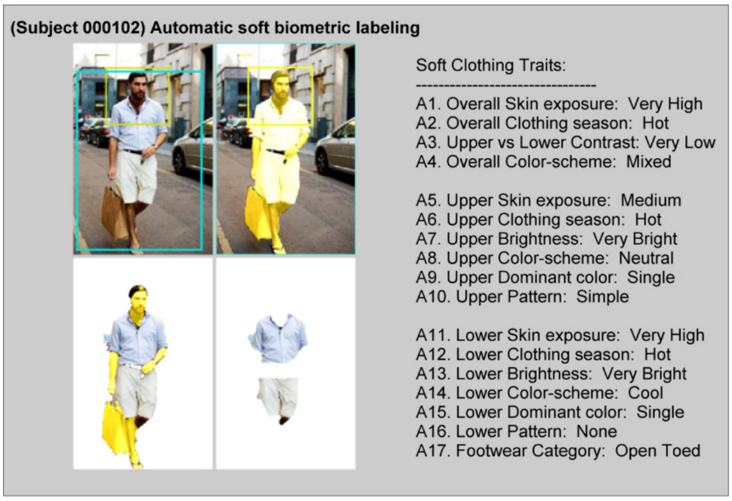
- **Applying Automatic Clothing Attribute Extraction on Images in the Wild**

Since we have already established in this thesis the work on automatic soft clothing attribute extraction and annotation, this triggers the next major milestone for applying our new automatic approaches to images in the wild for more realistic person retrieval, which can pave the way for novel and practical biometric applications. Thus we may consider the implementation and evaluation of these approaches on a wider selection of publically available databases e.g. the Clothing Attribute Dataset [1], VIPeR [13], APiS [5], or

PETA [121]. For instance, Fig. 8.1 shows two sample images belong to the Clothing Attribute Dataset [1], which we use as input images (captured in the wild) to test our automatic approach of clothing attribute extraction and annotation, described in Chapter 6. Moreover, it shows the resulting list automatically derived comprising 17 categorical soft labels describing clothing for each subject; more various examples are given in the appendix. These examples show how the proposed approach can be generalized to handle complex backgrounds and lighting so as to derive the clothing annotations for use in identification from such images.



Example (a): Automatic clothing annotation for subject 000096



Example (b): Automatic clothing annotation for subject 000102

Figure 8.1: Two examples of applying our automatic clothing attribute extraction and annotation on sample images (captured in the wild) from the Clothing Attribute Dataset

- **Exploration of Crowdsourcing and Its Influence on Clothing Trait Performance**

It is should be possible to carry on a study of the viability of soft clothing traits derived from crowdsourcing data annotations. This will allow comparing their performance with those traits derived from data collected from a limited number of invited users, and will investigate whether a much greater number of users would significantly influence the inter-annotator agreement or the data quality and performance. This study is considered as part of a Soton project aiming to create a large public database of soft biometrics containing soft clothing traits in addition to newly collected comparative soft body [122] and face [91] traits. This database will enable several research lines on capabilities, fusion, and integration of different soft biometrics, where this database can be used as benchmark for such research studies. Fig. 8.2 shows a crowdsourcing web-form, which we designed on CrowdFlower platform, to acquire comparative clothing labels for future research.

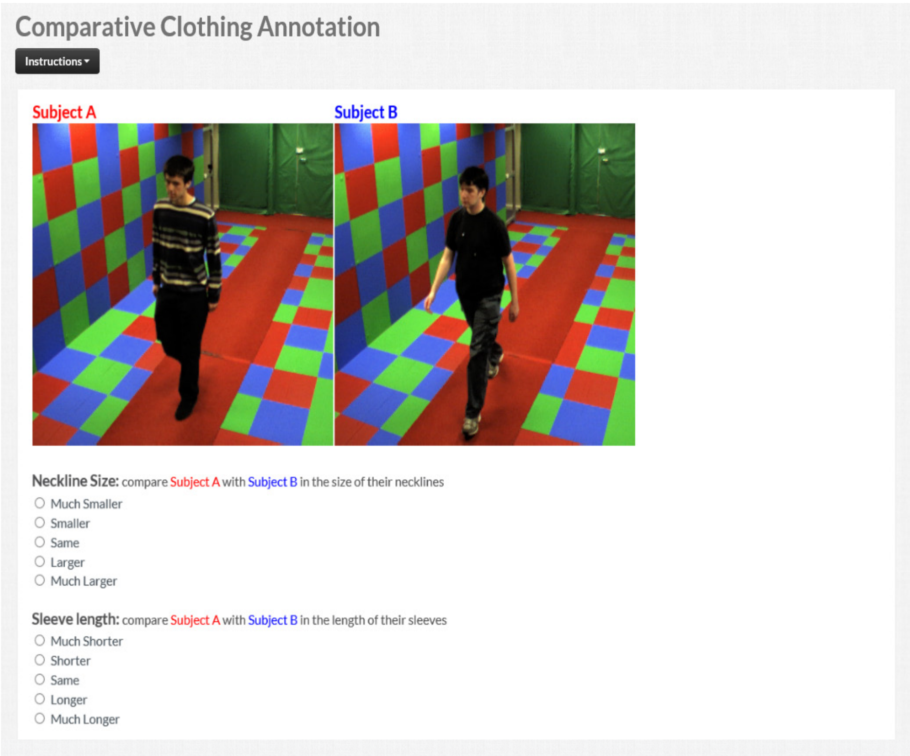


Figure 8.2: Comparative annotation web-form via CrowdFlower platform for crowdsourcing

- **Correlations and Dependency between Soft Clothing and Body Attributes**

A further extended analysis may consider investigations of improved integration between soft clothing and body biometrics for identity. This analysis can convey better understanding of the interaction and collaboration amongst soft clothing and body traits, leading to optimized feature selectin and fusion of these traits for more discriminative and functional soft biometric signatures. For this purpose, several investigations can be performed including trait correlations, statistical dependency, and mutual information. This does not only allow reliable feature selection and fusion but also enables predictability of clothing attributes from available body attributes and vice versa. Note that further exploration for possible pair correlations between categorical and comparative or manual and automatic soft clothing attributes will supply useful information to such future study; a variety of potential correlation matrices are given in the appendix. In this way we shall capitalise and extend this new approach.

References

[1] H. Chen, A. Gallagher, and B. Girod, "Describing clothing by semantic attributes," in *European Conference on Computer Vision (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 *Proc. British Machine Vision Conference (BMVC)*, 2015.

[3] R. Layne, T. M. Hospedales, and S. Gong, "Person Re-identification by Attributes," in *Proc. British Machine Vision Conference (BMVC)*, 2012.

[4] 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 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[5] J. Zhu, S. Liao, Z. Lei, D. Yi, and S. Z. Li, "Pedestrian Attribute Classification in Surveillance: Database and Evaluation," in *IEEE International Conference on Computer Vision (ICCV)*, 2013.

[6] L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. Van Gool, "Apparel classification with style," in *Asian Conference on Computer Vision (ACCV)*, 2013.

[7] S. Liu, J. Feng, Z. Song, T. Zhang, H. Lu, C. Xu, and S. Yan, "Hi, magic closet, tell me what to wear!," in *Proc. ACM International Conference on Multimedia (ACM MM12)*, 2012.

[8] D. A. Vaquero, R. S. Feris, D. Tran, L. Brown, A. Hampapur, and M. Turk, "Attribute-based people search in surveillance environments," in *IEEE Workshop on Applications of Computer Vision (WACV)*, 2009.

[9] A. Li, L. Liu, K. Wang, S. Liu, and S. Yan, "Clothing Attributes Assisted Person Re-identification," *IEEE Transactions on Circuits and Systems for Video Technology,* 2014.

[10] A. C. Gallagher and T. Chen, "Clothing cosegmentation for recognizing people," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[11] M. Toews and T. Arbel, "Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI),* vol. 31, pp. 1567-1581, 2009.

[12] A. Dantcheva and J.-L. Dugelay, "Frontal-to-side face re-identification based on hair, skin and clothes patches," in *8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, 2011, pp. 309-313.

[13] D. Gray, S. brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance*, 2007.

[14] X. Liang, L. Lin, W. Yang, P. Luo, J. Huang, and S. Yan, "Clothes Co-Parsing via Joint Image Segmentation and Labeling with Application to Clothing Retrieval," *IEEE Transactions on Multimedia,* 2016.

[15] J. Huang, R. S. Feris, Q. Chen, and S. Yan, "Cross-domain image retrieval with a dual attribute-aware ranking network," in *IEEE International Conference on Computer Vision (ICCV)*, 2015.

[16] S. Vittayakorn, K. Yamaguchi, A. C. Berg, and T. L. Berg, "Runway to realway: Visual analysis of fashion," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2015, pp. 951-958.

[17] M. Yang and K. Yu, "Real-time clothing recognition in surveillance videos," in *Proc. 18th IEEE International Conference on Image Processing (ICIP)*, 2011, pp. 2937-2940.

[18] G. Mu, D. Yuning, A. Haizhou, and L. Shihong, "A hybrid approach to pedestrian clothing color attribute extraction," in *14th IAPR International Conference on Machine Vision Applications (MVA)*, 2015, pp. 81-84.

[19] 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.

[20] M. Halstead, S. Denman, S. Sridharan, and C. Fookes, "Locating people in video from semantic descriptions: A new database and approach," in *Proc. International Conference on Pattern Recognition (ICPR)*, 2014.

[21] Z. Shi, T. M. Hospedales, and T. Xiang, "Transferring a semantic representation for person re-identification and search," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[22] 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 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5315-5324.

[23] A. Kovashka, D. Parikh, and K. Grauman, "Whittlesearch: Image search with relative attribute feedback," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2973-2980.

[24] E. S. Jaha and L. Ghouti, "Color face recognition using quaternion PCA," in *4th International Conference on Imaging for Crime Detection and Prevention (ICDP)*, 2011.

[25] N. Poh, R. Wong, J. Kittler, and F. Roli, "Challenges and research directions for adaptive biometric recognition systems," in *Advances in Biometrics*, ed: Springer, 2009, pp. 753-764.

[26] A. K. Jain, S. C. Dass, and K. Nandakumar, "Can soft biometric traits assist user recognition?," in *Proc. SPIE 5404, Biometric Technology for Human Identification*, 2004, pp. 561-572.

[27] D. Reid, M. Nixon, and S. Stevenage, "Soft Biometrics; Human Identification using Comparative Descriptions," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI),* vol. 36, pp. 1216-1228, 2014.

[28] X. Chen and B. Bhanu, "Soft Biometrics Integrated Multi-target Tracking," in *Proc. International Conference on Pattern Recognition (ICPR)*, 2014.

[29] S. Samangooei, B. Guo, and M. S. Nixon, "The use of semantic human description as a soft biometric," in *Proc. 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2008.

[30] 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.

[31] D. Parikh and K. Grauman, "Relative attributes," in *IEEE International Conference on Computer Vision (ICCV)*, 2011.

[32] A. Nambiar, A. Bernardino, and J. Nascimento, "Shape Context for soft biometrics in person re-identification and database retrieval," *Pattern Recognition Letters,* vol. 68, pp. 297-305, 2015.

[33] K. Niinuma, U. Park, and A. K. Jain, "Soft biometric traits for continuous user authentication," *IEEE Transactions on Information Forensics and Security (TFIS),* vol. 5, pp. 771-780, 2010.

[34] D. A. Reid and M. S. Nixon, "Imputing human descriptions in semantic biometrics," in *Proc. 2nd ACM workshop on Multimedia in forensics, security and intelligence*, 2010, pp. 25-30.

[35] D. A. Reid and M. S. Nixon, "Using comparative human descriptions for soft biometrics," in *International Joint Conference on Biometrics (IJCB)*, 2011.

[36] M. Ao, D. Yi, Z. Lei, and S. Z. Li, "Face recognition at a distance: System issues," in *Handbook of Remote Biometrics*, ed: Springer, 2009, pp. 155-167.

[37] O. A. Arigbabu, S. M. S. Ahmad, W. A. W. Adnan, and S. Yussof, "Integration of multiple soft biometrics for human identification," *Pattern Recognition Letters,* vol. 68, pp. 278-287, 2015.

[38] U. Park and A. K. Jain, "Face matching and retrieval using soft biometrics," *IEEE Transactions on Information Forensics and Security (TFIS),* vol. 5, pp. 406-415, 2010.

[39] 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.

[40] P. Tome, J. Fierrez, R. Vera-Rodriguez, and M. Nixon, "Soft Biometrics and Their Application in Person Recognition at a Distance," *IEEE Transactions on Information Forensics and Security (TFIS),* 2014.

[41] D. A. Reid, M. S. Nixon, and S. V. Stevenage, "Identifying humans using comparative descriptions," in *Proc. 4th International Conference on Imaging for Crime Detection and Prevention (ICDP)*, 2011.

[42] J. Thornton, J. Baran-Gale, D. Butler, M. Chan, and H. Zwahlen, "Person attribute search for large-area video surveillance," in *IEEE International Conference on Technologies for Homeland Security (HST)*, 2011, pp. 55-61.

[43] D. Adjeroh, C. Deng, M. Piccirilli, and A. Ross, "Predictability and correlation in human metrology," in *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2010.

[44] B. Siddiquie, R. S. Feris, and L. S. Davis, "Image ranking and retrieval based on multi-attribute queries," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 801-808.

[45] X. Zhou, B. Bhanu, and J. Han, "Human recognition at a distance in video by integrating face profile and gait," in *Face Biometrics for Personal Identification*, ed: Springer, 2007, pp. 165-181.

[46] D. A. Reid and M. S. Nixon, "Human identification using facial comparative descriptions," in *International Conference on Biometrics (ICB)*, 2013.

[47] A. Dantcheva, J.-L. Dugelay, and P. Elia, "Person recognition using a bag of facial soft biometrics (BoFSB)," in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, 2010, pp. 511-516.

[48] A. K. Jain and U. Park, "Facial marks: Soft biometric for face recognition," in *Proc. 16th IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 37-40.

[49] C. Deng, C. Cunjian, M. Piccirilli, D. Adjeroh, T. Bourlai, and A. Ross, "Can facial metrology predict gender?," in *International Joint Conference on Biometrics (IJCB)*, 2011.

[50] L. Best-Rowden, S. Bisht, J. C. Klontz, and A. K. Jain, "Unconstrained Face Recognition: Establishing Baseline Human Performance via Crowdsourcing," *International Joint Conference on Biometrics (IJCB),* 2014.

[51] B. F. Klare, S. Klum, J. C. Klontz, E. Taborsky, T. Akgul, and A. K. Jain, "Suspect Identification Based on Descriptive Facial Attributes," in *International Joint Conference on Biometrics (IJCB)*, 2014.

[52] O. A. Arigbabu, S. M. S. Ahmad, W. A. W. Adnan, and S. Yussof, "Recent advances in facial soft biometrics," *The Visual Computer,* 2014.

[53] H. Zhang, J. R. Beveridge, B. A. Draper, and P. J. Phillips, "On the effectiveness of soft biometrics for increasing face verification rates," *Computer Vision and Image Understanding,* vol. 137, pp. 50-62, 2015.

[54] D. Mery and K. Bowyer, "Automatic facial attribute analysis via adaptive sparse representation of random patches," *Pattern Recognition Letters,* vol. 68, pp. 260-269, 2015.

[55] A. K. Jain, K. Nandakumar, X. Lu, and U. Park, "Integrating faces, fingerprints, and soft biometric traits for user recognition," in *Biometric Authentication*, ed: Springer, 2004, pp. 259-269.

[56] S. Denman, A. Bialkowski, C. Fookes, and S. Sridharan, "Determining operational measures from multi-camera surveillance systems using soft biometrics," in *Proc. 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, 2011, pp. 462-467.

[57] S. Denman, C. Fookes, A. Bialkowski, and S. Sridharan, "Soft-Biometrics: Unconstrained Authentication in a Surveillance Environment," in *Digital Image Computing: Techniques and Applications (DICTA '09)*, 2009, pp. 196-203.

[58] Y.-F. Wang, E. Y. Chang, and K. P. Cheng, "A video analysis framework for soft biometry security surveillance," in *Proc. 3rd ACM international workshop on Video surveillance & sensor networks*, 2005, pp. 71-78.

[59] L. Zhang, D. V. Kalashnikov, S. Mehrotra, and R. Vaisenberg, "Context-based person identification framework for smart video surveillance," *Machine Vision and Applications,* vol. 25, pp. 1711-1725, 2014.

[60] L. Zhang, D. V. Kalashnikov, and S. Mehrotra, "Context Assisted Person Identification for Images and Videos," *Handbook of Pattern Recognition and Computer Vision,* 2015.

[61] A. Bialkowski, S. Denman, P. Lucey, S. Sridharan, and C. B. Fookes, "A database for person re-identification in multi-camera surveillance networks," in *Proc. International Conference on Digital Image Computing Techniques and Applications (DICTA 12)*, 2012.

[62] G. Guo, "Human age estimation and sex classification," in *Video Analytics for Business Intelligence*, ed: Springer, 2012, pp. 101-131.

[63] R. Vezzani, D. Baltieri, and R. Cucchiara, "People reidentification in surveillance and forensics: A survey," *ACM Computing Surveys (CSUR),* vol. 46, p. 29, 2013.

[64] A. Vrij, H. Pannell, and J. Ost, "The influence of social pressure and black clothing on crime judgements," *Psychology, Crime & Law,* vol. 11, pp. 265-274, 2005.

[65] X. Chao, M. J. Huiskes, T. Gritti, and C. Ciuhu, "A framework for robust feature selection for real-time fashion style recommendation," in *Proc. 1st international workshop on Interactive multimedia for consumer electronics*, 2009, pp. 35-42.

[66] G. A. Cushen and M. S. Nixon, "Real-time semantic clothing segmentation," in *Advances in Visual Computing*, ed: Springer, 2012, pp. 272-281.

[67] W. Zhang, B. Begole, M. Chu, J. Liu, and N. Yee, "Real-time clothes comparison based on multi-view vision," in *Proc. 2nd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)* 2008.

[68] T. Avraham and M. Lindenbaum, "Learning Appearance Transfer for Person Re-identification," in *Person Re-Identification*, ed: Springer, 2014, pp. 231-246.

[69] M. Demirkus, K. Garg, and S. Guler, "Automated person categorization for video surveillance using soft biometrics," in *Proc. SPIE 7667, Biometric Technology for Human Identification* 2010, pp. 76670P-1.

[70] S. Sunderrajan and B. Manjunath, "Context-Aware Hypergraph Modeling for Re-identification and Summarization," *IEEE Transactions on Multimedia,* vol. 18, pp. 51-63, 2016.

[71] X. Chen, L. An, and B. Bhanu, "Soft-Biometrics and Reference Set Integrated Model for Tracking Across Cameras," in *Distributed Embedded Smart Cameras*, ed New York: Springer, 2014, pp. 211-230.

[72] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher, "Appearance-based person reidentification in camera networks: problem overview and current approaches," *Journal of Ambient Intelligence and Humanized Computing,* vol. 2, pp. 127-151, 2011.

[73] S. Cai, J. Wang, and L. Quan, "How Fashion Talks: Clothing-Region-Based Gender Recognition," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, ed: Springer, 2014, pp. 515-523.

[74] H. Wang, X. Bao, R. R. Choudhury, and S. Nelakuditi, "InSight: recognizing humans without face recognition," in *Proc. 14th Workshop on Mobile Computing Systems and Applications*, 2013.

[75] M. S. Nixon, P. Correia, K. Nasrollahi, T. Moeslund, A. Hadid, and M. Tistarelli, "On Soft Biometrics," *Pattern Recognition Letters,* 2015.

[76] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? A survey on soft biometrics," *IEEE Transactions on Information Forensics and Security (TIFS),* 2015.

[77] A. Schumann and E. Monari, "A soft-biometrics dataset for person tracking and re-identification," in *11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2014, pp. 193-198.

[78] A. Li, L. Liu, and S. Yan, "Person Re-identification by Attribute-Assisted Clothes Appearance," in *Person Re-Identification*, ed: Springer, 2014, pp. 119-138.

[79] R. Layne, T. M. Hospedales, and S. Gong, "Attributes-Based Re-identification," in *Person Re-Identification*, ed: Springer, 2014, pp. 93-117.

[80] R. Layne, T. M. Hospedales, and S. Gong, "Towards person identification and re-identification with attributes," in *European Conference on Computer Vision (ECCV)*, 2012, pp. 402-412.

[81] T. Wang, S. Gong, X. Zhu, and S. Wang, "Person Re-Identification by Discriminative Selection in Video Ranking," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI),* 2016.

[82] 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.

[83] J. Shutler, M. Grant, M. S. Nixon, and J. N. Carter, "On a large sequence-based human gait database," in *Proc. Recent Advances in Soft Computing (RASC02)*, 2002.

[84] G. B. Chapman and E. J. Johnson, "Incorporating the irrelevant: Anchors in judgments of belief and value," *Heuristics and biases: The psychology of intuitive judgment,* pp. 120-138, 2002.

[85] H. D. Ellis, "Practical aspects of face memory," in *Eyewitness testimony: Psychological perspectives*, ed: Cambridge University Press, 1984, pp. 12-37.

[86] R. H. Flin and J. W. Shepherd, "Tall stories: Eyewitnesses' ability to estimate height and weight characteristics," *Human Learning: Journal of Practical Research & Applications,* 1986.

[87] A. J. O'Toole, "Psychological and neural perspectives on human face recognition," in *Handbook of face recognition*, ed: Springer, 2005, pp. 349-369.

[88] R. C. Lindsay, R. Martin, and L. Webber, "Default values in eyewitness descriptions: A problem for the match-to-description lineup foil selection strategy," *Law and Human Behavior,* vol. 18, p. 527, 1994.

[89] T. Matthews, M. Nixon, and M. Niranjan, "Enriching texture analysis with semantic data," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 1248-1255.

[90] T. Joachims, "Optimizing search engines using clickthrough data," in *Proc. ACM Special Interest Group on Knowledge Discovery and Data Mining (ACM SIGKDD)*, 2002, pp. 133-142.

[91] N. Almudhahka, M. Nixon, and J. Hare, "Human face identification via comparative soft biometrics," in *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, 2016.

[92] S. Samangooei and M. S. Nixon, "On Semantic Soft-Biometric Labels," in *Biometric Authentication*, ed: Springer, 2014, pp. 3-15.

[93] A. Field, *Discovering statistics using SPSS*, 3rd ed.: Sage publications, 2009.

[94] J. Pohjalainen, O. Räsänen, and S. Kadioglu, "Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits," *Computer Speech & Language,* vol. 29, pp. 145-171, 2015.

[95] R. Kirkham, A. Khan, S. Bhattacharya, N. Hammerla, S. Mellor, D. Roggen, and T. Ploetz, "Automatic correction of annotation boundaries in activity datasets by class separation maximization," in *Proc. ACM conference on Pervasive and ubiquitous computing adjunct publication*, Zurich, Switzerland, 2013, pp. 673-678.

[96] A. Peldszus and M. Stede, "Ranking the annotators: An agreement study on argumentation structure," in *Proc. 7th linguistic annotation workshop and interoperability with discourse*, 2013, pp. 196-204.

[97] V. Bhardwaj, R. J. Passonneau, A. Salleb-Aouissi, and N. Ide, "Anveshan: a framework for analysis of multiple annotators' labeling behavior," in *Proc. 4th Linguistic Annotation Workshop*, 2010, pp. 47-55.

[98] Z. Theodosiou, C. Kasapi, and N. Tsapatsoulis, "Semantic Gap between People: An Experimental Investigation Based on Image Annotation," in *7th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, 2012, pp. 73-77.

[99] F. K. Khattak and A. Salleb-Aouissi, "Improving Crowd Labeling through Expert Evaluation," in *AAAI Spring Symposium: Wisdom of the Crowd*, 2012.

[100] R. J. Passonneau and B. Carpenter, "The benefits of a model of annotation," in *Proc. 7th Linguistic Annotation Workshop and Interoperability with Discourse*, 2013, pp. 187-195.

[101] J. Cohen, "A coeffiecient of agreement for nominal scales," *Educational and Psychological Measurement,* vol. 20, pp. 37–46, 1960.

[102] J. Carletta, "Assessing agreement on classification tasks: the kappa statistic," *Computational linguistics,* vol. 22, pp. 249-254, 1996.

[103] I. Jena, R. A. Bhat, S. Jain, and D. M. Sharma, "Animacy Annotation in the Hindi Treebank," in *7th Linguistic Annotation Workshop & Interoperability with Discourse (LAW-VII & ID)*, 2013.

[104] S. Rosset, C. Grouin, T. Lavergne, M. Ben Jannet, J. Leixa, O. Galibert, and P. Zweigenbaum "Automatic Named Entity Pre-Annotation for Out-of-Domain Human Annotation," *7th Linguistic Annotation Workshop & Interoperability with Discourse (LAW-VII & ID),* p. 168, 2013.

[105] D. Reid, S. Samangooei, C. Chen, M. Nixon, and A. Ross, "Soft biometrics for surveillance: an overview," in *Handbook of Statistics: Machine Learning: Theory and Applications*, ed: Elsevier, 2013, pp. 327-352.

[106] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI),* vol. 32, pp. 1627-1645, 2010.

[107] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001.

[108] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, "Progressive search space reduction for human pose estimation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1-8.

[109] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *ACM Transactions on Graphics (TOG),* vol. 23, pp. 309-314, 2004.

[110] D. A. Winter, *The Biomechanics and Motor Control of Human Movement*, 2nd ed.: John Wiley & Sons, 1990.

[111] C. O. Conaire, N. E. O'Connor, and A. F. Smeaton, "Detector adaptation by maximising agreement between independent data sources," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[112] D. E. Ilea and P. F. Whelan, "Color image segmentation using a spatial k-means clustering algorithm," in *Irish Machine Vision and Image Processing Conference*, 2006.

[113] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI),* vol. 24, pp. 971-987, 2002.

[114] L. Yang, D. Zhao, X. Wu, H. Li, and J. Zhai, "An improved Prewitt algorithm for edge detection based on noised image," in *4th International Congress on Image and Signal Processing (CISP)*, 2011, pp. 1197-1200.

[115] B. Guo and M. S. Nixon, "Gait feature subset selection by mutual information," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans,* vol. 39, pp. 36-46, 2009.

[116] B. Guo, R. I. Damper, S. R. Gunn, and J. D. Nelson, "A fast separability-based feature-selection method for high-dimensional remotely sensed image classification," *Pattern Recognition,* vol. 41, pp. 1653-1662, 2008.

[117] C. Boon Ng, Y. Haur Tay, and B. M. Goi, "Vision-based Human Gender Recognition: A Survey," in *Proc. Pacific Rim International Conference Artificial Intelligence: Trends Artificial Intelligence*, 2012, pp. 335-346.

[118] 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 Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* , 2012, pp. 130-137.

[119] L. Bourdev, S. Maji, and J. Malik, "Describing people: A poselet-based approach to attribute classification," in *IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 1543-1550.

[120] L. Cao, M. Dikmen, Y. Fu, and T. S. Huang, "Gender recognition from body," in *Proc. 16th ACM international conference on Multimedia*, 2008, pp. 725-728.

[121] Y. Deng, P. Luo, C. C. Loy, and X. Tang, "Pedestrian attribute recognition at far distance," in *Proc. of the ACM International Conference on Multimedia*, 2014, pp. 789-792.

[122] D. Martinho-Corbishley, M. S. Nixon, and J. N. Carter, "Soft biometric recognition from comparative crowdsourced annotations," in *I6th International Conference on maging for Crime Prevention and Detection (ICDP-15)*, 2015.

Appendix A

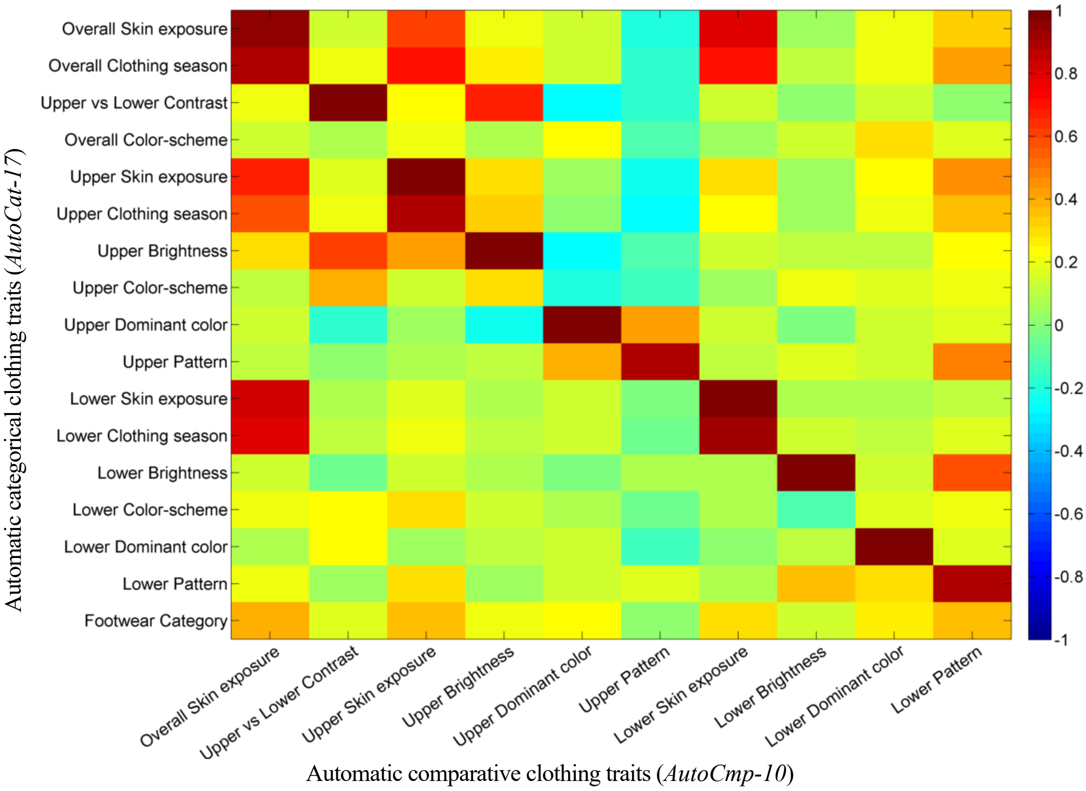


Figure 0.1: Correlations between *AutoCat-17* and *AutoCmp-10* soft clothing traits

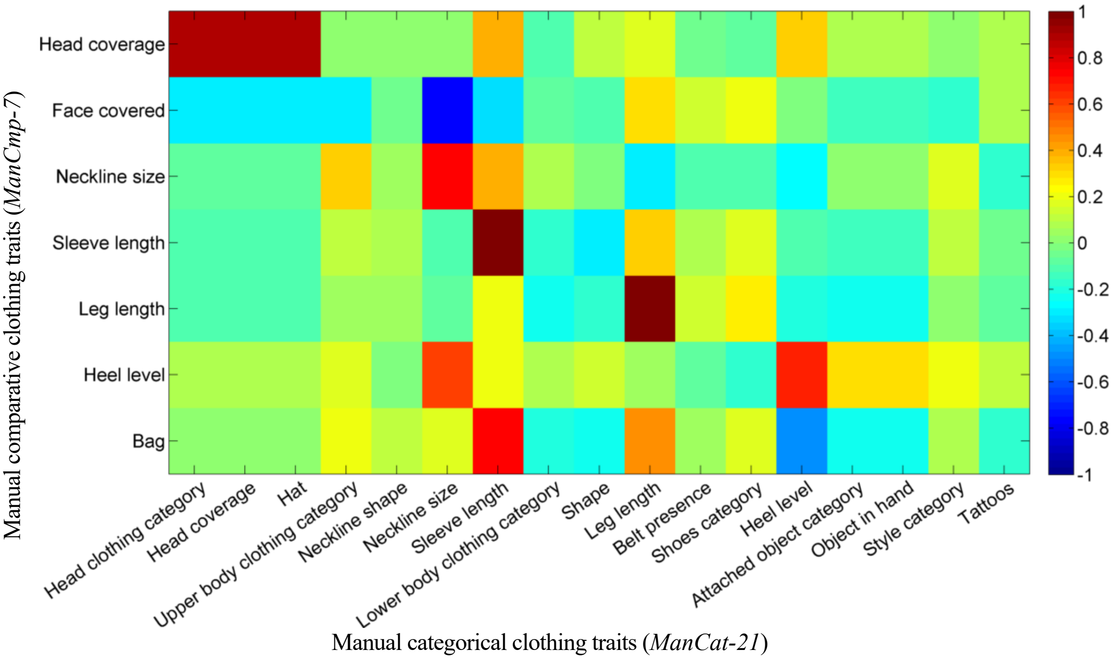


Figure 0.2: Correlations between significant traits of *ManCat-21* and *ManCmp-7* soft clothing traits

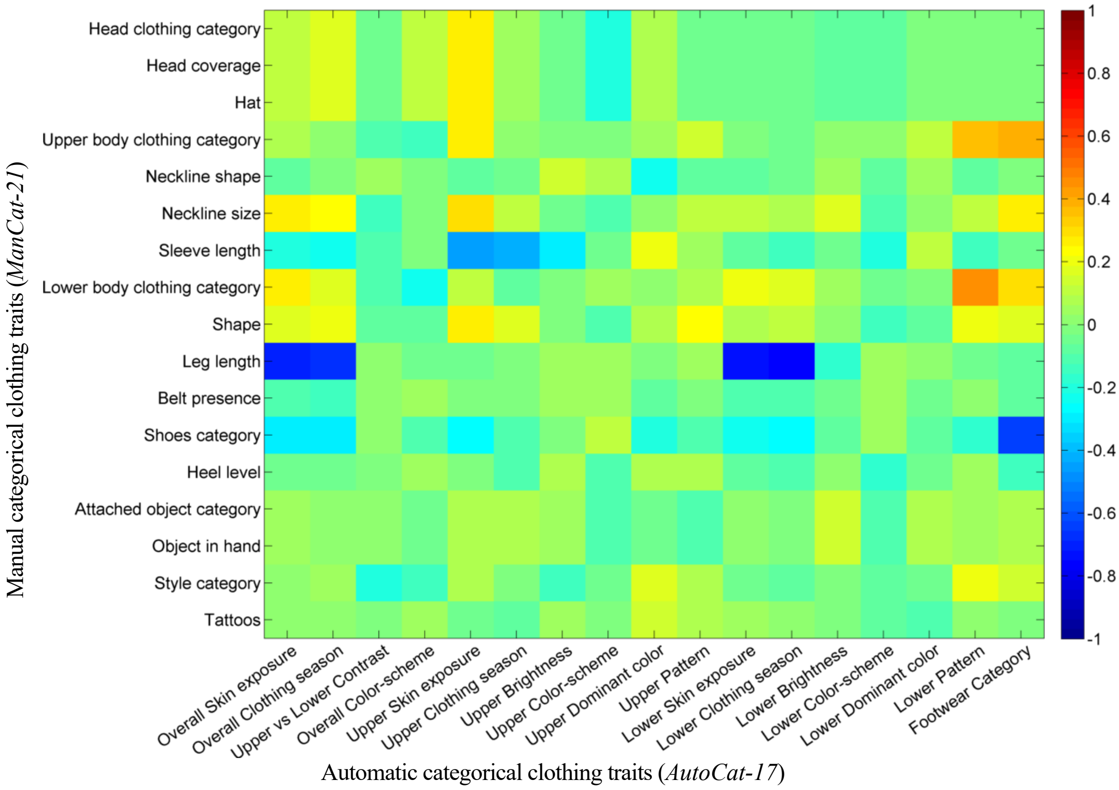


Figure 0.3: Correlations between significant traits of *ManCat-21* and *AutoCat-17* soft clothing traits

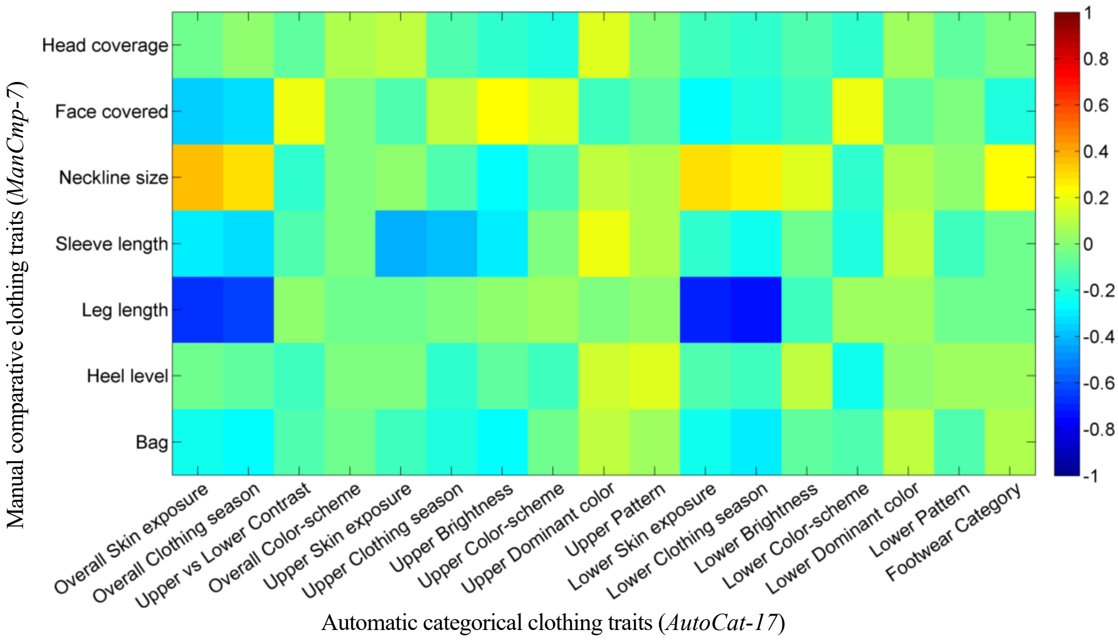


Figure 0.4: Correlations between *ManCmp-7* and *AutoCat-17* soft clothing traits

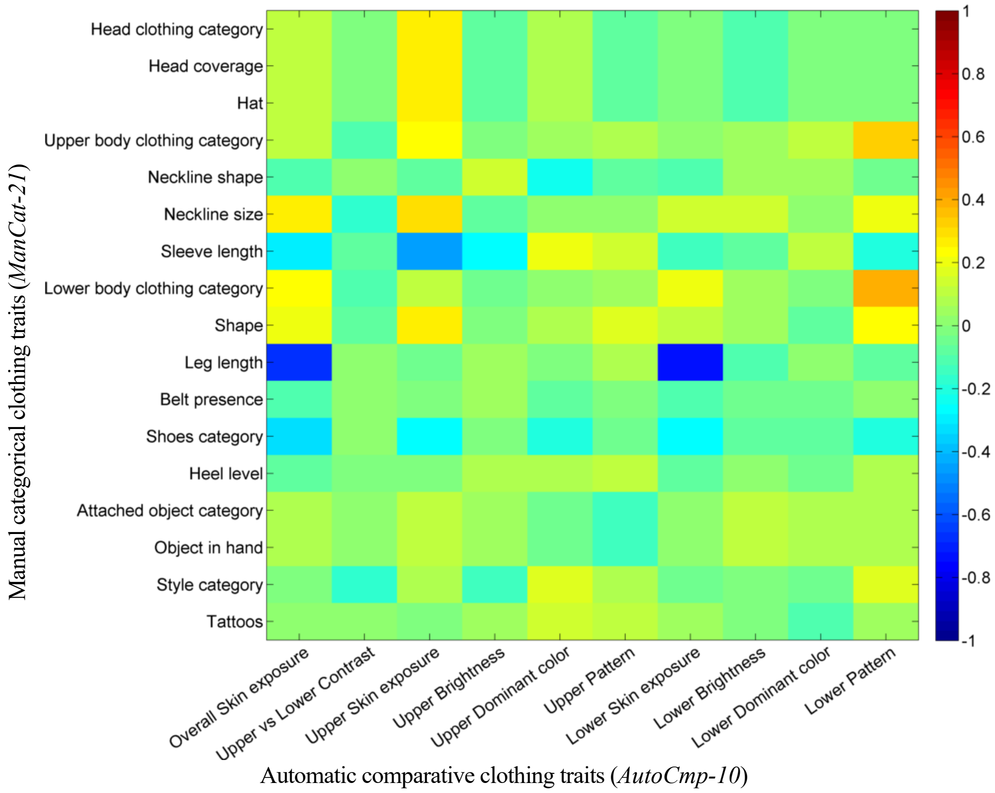


Figure 0.5: Correlations between significant traits of *ManCat-21* and *AutoCmp-10* soft clothing traits

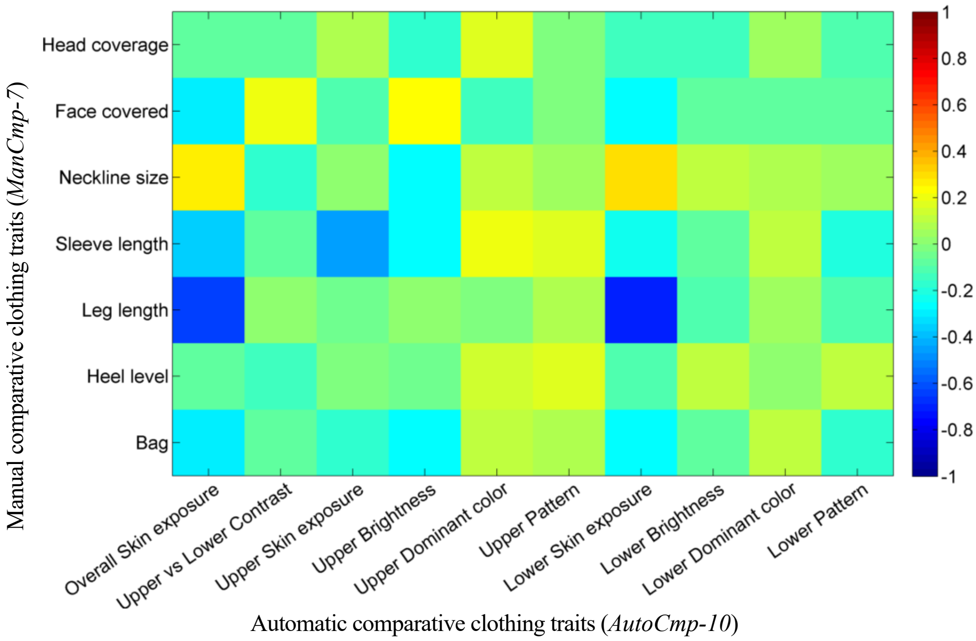
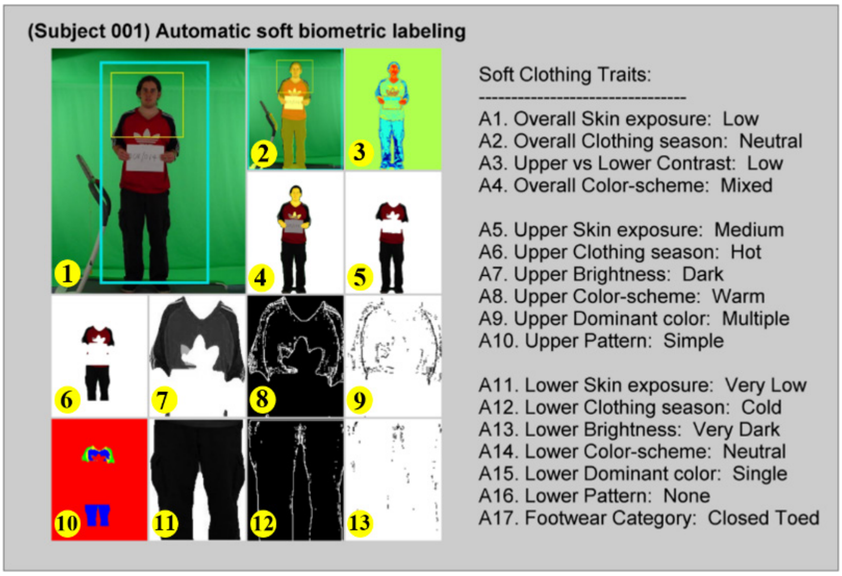
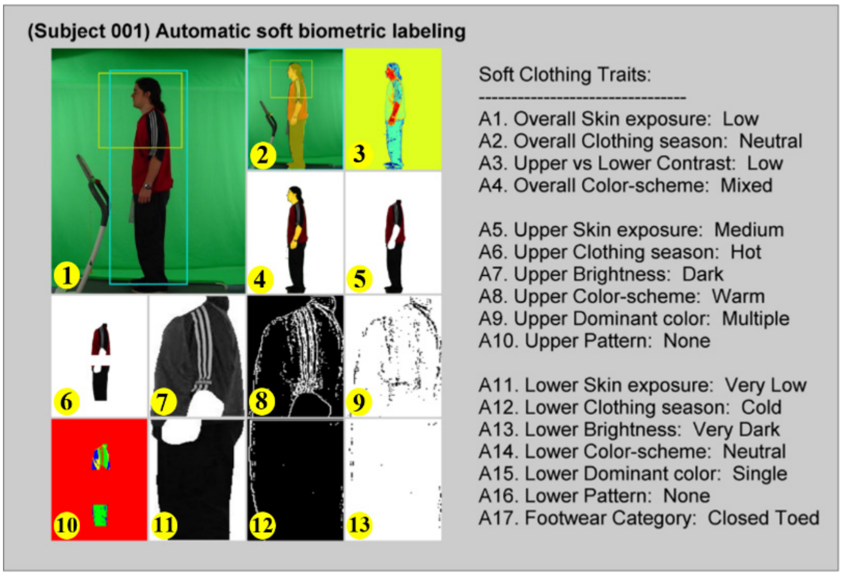


Figure 0.6: Correlations between *ManCmp-7* and *AutoCmp-10* soft clothing traits



(i) Automatic clothing annotation for a front-view subject image

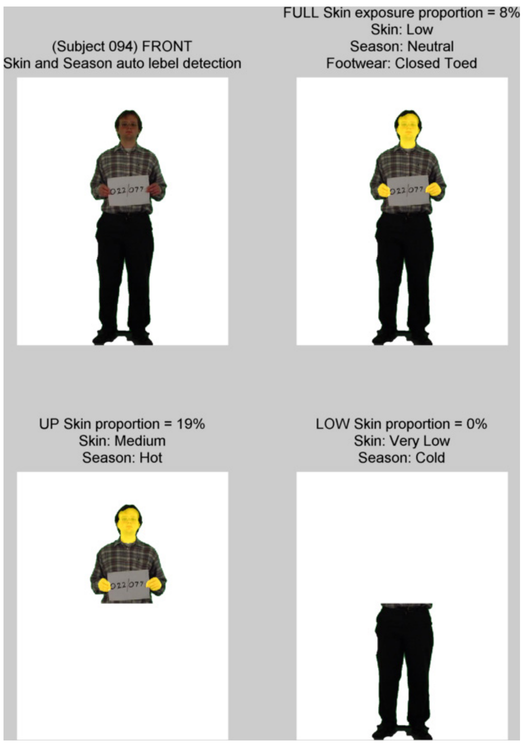


(ii) Automatic clothing annotation for a side-view subject image

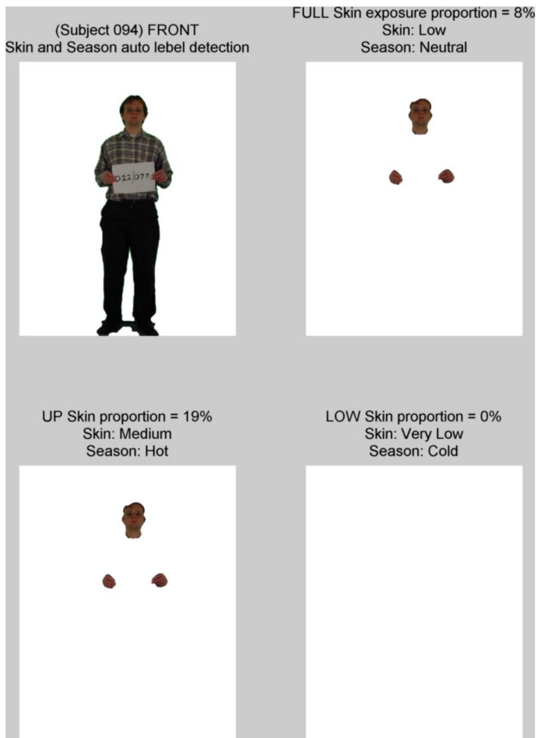
Figure 0.7: A summary screen of a set of all 17 categorical soft clothing attributes automatically extracted and labelled to describe (i) front- and (ii) side-viewpoint subject images. Sub-images demonstrate some details of the applied automatic vision-based image processing as follows: *preprocessing*, (1) face/body detection, and (2) subject segmentation; *analysing appearance*, (3-4) skin detection and (5-6) upper/lower clothing segmentation; *analysing colours*, (10) dominant colour quantization and (6, 7, and 11) hues/brightness/contrast analysis; and *texture analysis*, (8 and 12) ULBP representation and (9 and 13) complement binary representation of complexity

a-1. Model-based a-2. RGB-based a-3. HSV-based a-4. YCbCr-based a-5. Final detected skin

(a) Compound skin detectors



(b) Clothing season attributes



(c) Skin exposure attributes

Figure 0.8: Part of automatic appearance analysis and annotation

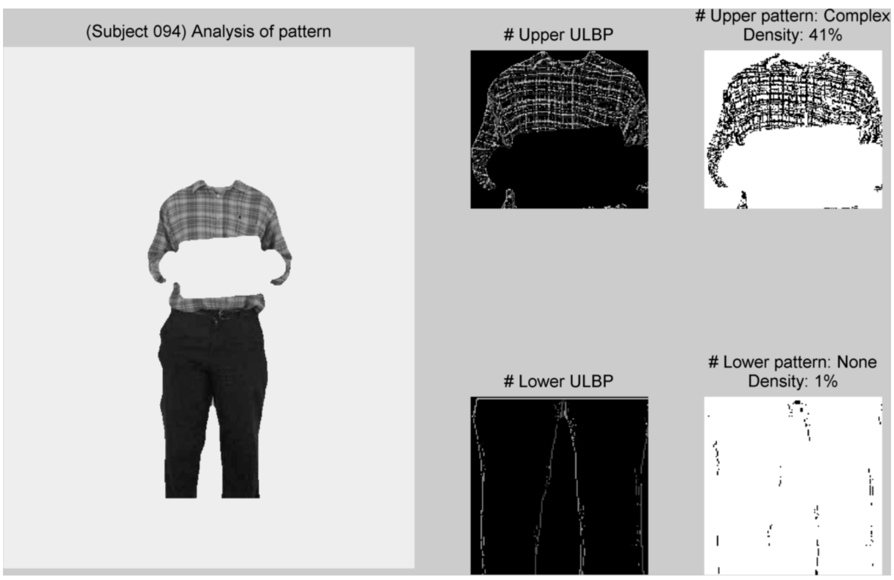
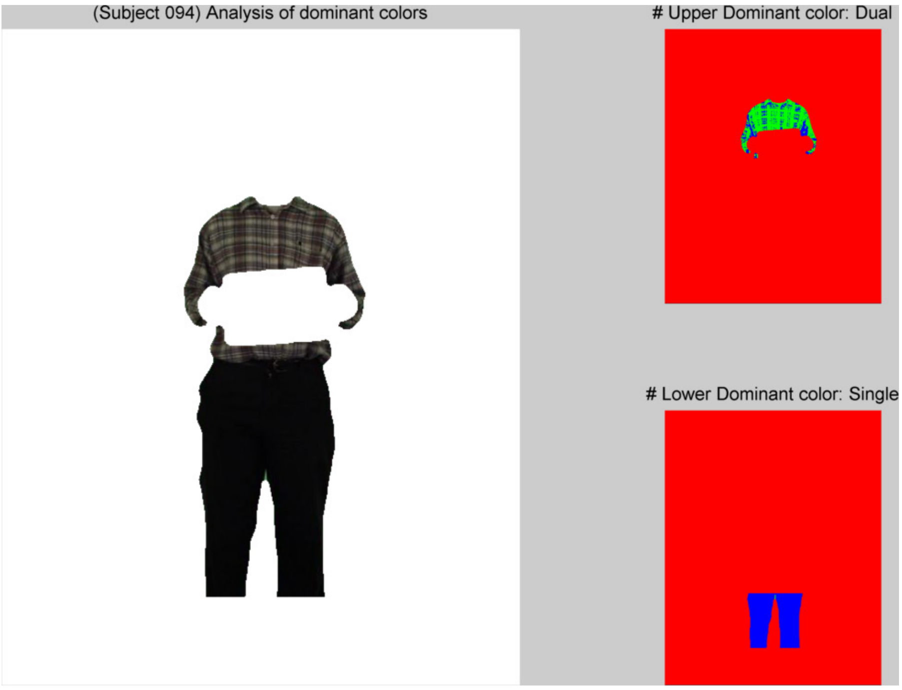


Figure 0.9: A part of automatic texture analysis and annotation for pattern complexity attributes

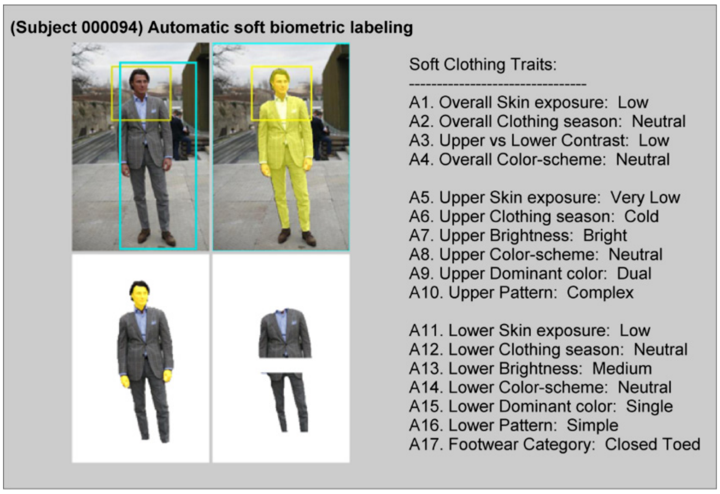
(a) Brightness and contrast attributes

(a) Colour-scheme attributes

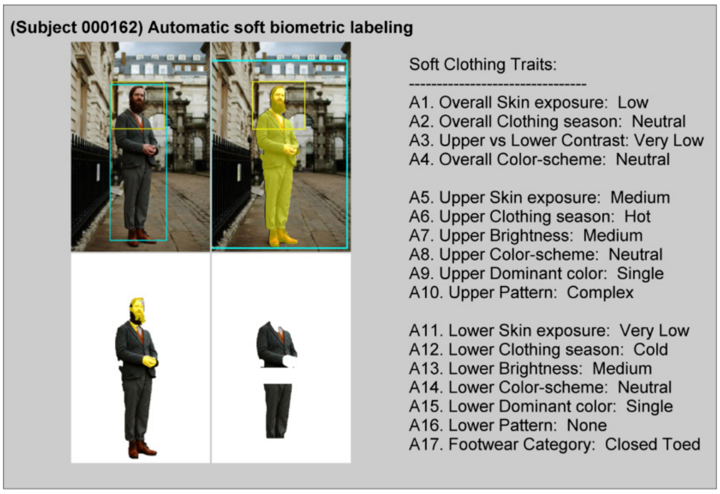


(a) Dominant colour attributes

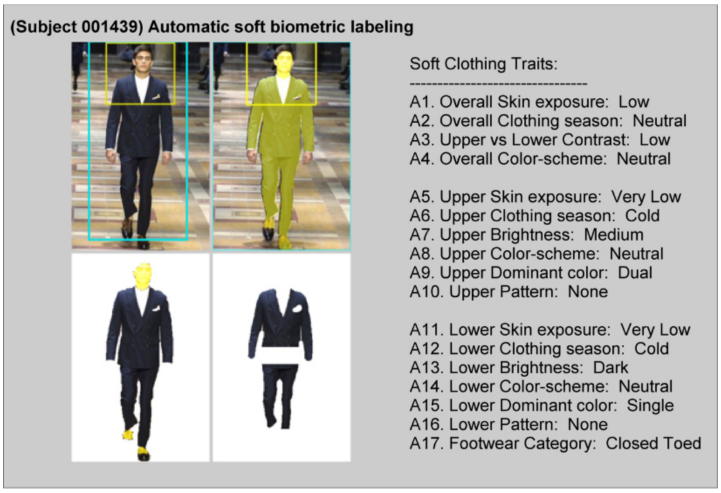
Figure 0.10: Part of automatic colour analysis and annotation



Example (a): Automatic clothing annotation for subject 000094

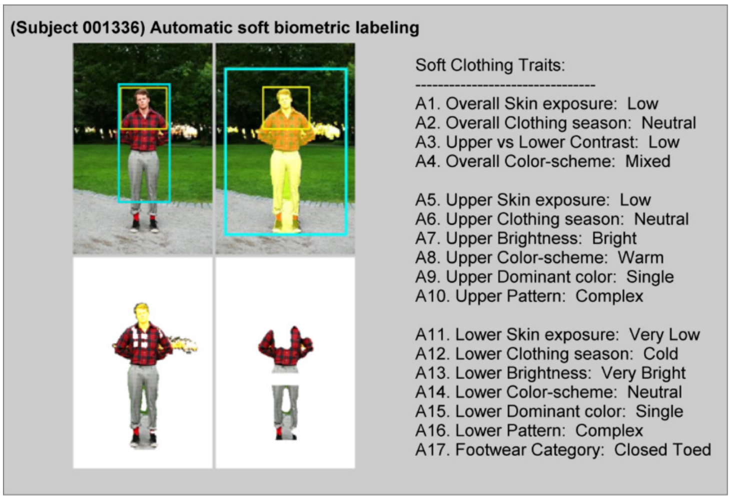


Example (b): Automatic clothing annotation for subject 000162 (semi-side viewpoint)

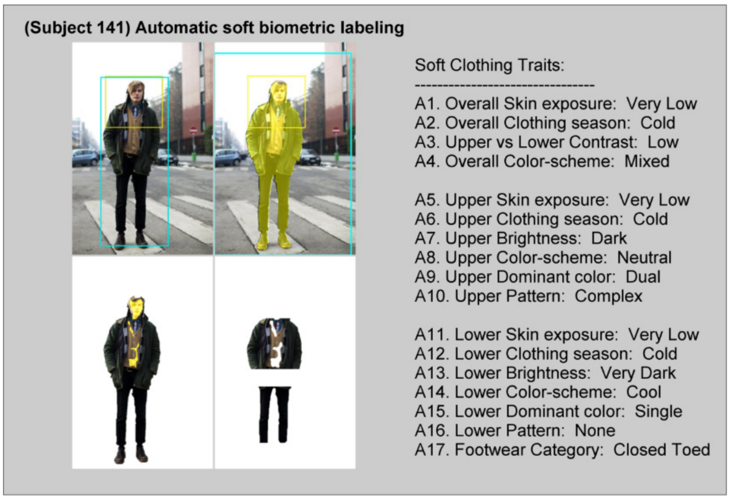


Example (c): Automatic clothing annotation for subject 001439

Figure 0.11: More examples of applying our automatic clothing attribute extraction and annotation on sample images (captured in the wild) from the Clothing Attribute Dataset



Example (a): Automatic clothing annotation for subject 001336



Example (b): Automatic clothing annotation for subject 000141

Figure 0.12: Extra examples of applying our automatic clothing attribute extraction and annotation on sample images (captured in the wild) from the Clothing Attribute Dataset