Shape-based Person Re-identification

A Thesis Submitted to the University of Kent

For the Degree of Doctor of Philosophy

In Electronic Engineering

By

Rawabi Alsedais

August 2019

Abstract

The increasing demand for public security, including forensic security, has resulted in a substantial growth in the presence of surveillance camera networks (i.e., closed-circuit televisions, or CCTVs) in public areas. Significant improvements in the computer vision and machine learning fields have advanced the traditional surveillance camera network system (i.e., monitored by people) towards an intelligent surveillance system involving automated person detection, person tracking, activity recognition, and person re-identification. The field of person re-identification has recently received much attention from computer vision researchers. Appearance model-based features, which are detection features that are built based on elements of the subject's appearance, such as texture, colour, and clothes, are used in person re-identification. However, using the body shape (as one of the appearance model-based features) as a signature for person re-identification is an area of research still open for examination.

This thesis presents the methodology, implementation, and experimental framework of a shape-based person re-identification system. The proposed system segments the human silhouette into eight different parts: *Body*, *Head & Neck*, *Shoulders*, *Middle*, *Lower*, *Upper Quarter*, *Upper Half*, *Torso,* and *Lower Half*. These segmentations are built based on anthropometry studies. This system exploits the shape descriptor information of these segments to build a subject-unique signature for person re-identification using a Generic Fourier Descriptor (GFD). The discrimination level of shape-based signatures are assessed by classifying them using image-based and video-based approaches. The image-based system classifies the signatures on a frame-by-frame basis using Linear Discriminant Analysis (LDA), which evaluates the feasibility of re-identifying subjects based on their shape static feature. The video-based approach exploits the signatures of the entire sequence (i.e., multiple frames) to re-identify subjects based on their dynamic features that occur within a collection of frames using Dynamic Time Wrapping (DTW). Comprehensive system outcomes for image-based and video-based systems are analysed by comparing the performance of both systems for each segment individually. Finally, a rank list fusion method, which combines the image-based generated rank lists so that the lists generated by all frames in each sequence are replaced by one rank list for the entire sequence, is implemented for performance enhancement.

Extensive experiments were conducted using publicly available dataset to evaluate the proposed shape-based person re-identification. In scenarios where a subject who maintains the same appearance is identified and re-identified from the same angle, the image-based and video-based approaches were found to outperform a number of state-of-art systems. In situations where the subject is identified and re-identified from different viewing angles (inter-view) and with a change in appearance (cross-scenario), the results reflected a comparable performance. The results of the rank list fusion implementation indicate superior performance enhancement in all situations, including the inter-view and cross-scenario.

Acknowledgments

I would like to express my sincere gratitude to my supervisor Richard Guest for the advice, encouragement and continuous supports received from him during my PhD study. Richard Guest gave me this precious opportunity to pursue PhD under his supervision. His guidance helped me in all the time of research and writing of this thesis.

I express grateful thank to the University of Kent. I would also like to thank King Faisal University in the Kingdom of Saudi Arabia for providing the funding which allowed me to undertake this study.

A special thanks to my family. Words cannot express how grateful I am to my mother, father (Modhi & Abdulaziz), sisters Jawharah, Aisha, Reem, Yasmin and Kenaz, and brothers Abdullah and Abdurrahman for always believing in me and encouraging me to follow my dreams. Your prayer for me was what sustained me thus far. I would also like to thank my mother-in law Modhi for her continuous support and endless love.

I must express my very profound gratitude to my friend Nawal and her family for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. I am indebted to all my friends in Canterbury who were always so helpful in numerous ways.

And finally, I thank with love my husband Abdullah, who has been by my side throughout this PhD, living every single minute of it, and being always my support in the moments when there was no one to answer my queries. And to my children Modhi & Ibrahim for being such good kids and making it possible for me to complete what I started. This thesis is dedicated to them.

Content

Lists of Tables

List of Figures

Lists of Abbreviations

Abbreviation	Meaning
1D	One Dimension
2D	Two Dimension
CC	Contour Coordinate
CCTV	Closed-circuit television
CMC	Cumulative matching characteristics
CNN	Conventional Neural Networks
DFT	Discrete Fourier transform
DTW	Dynamic Time Wrapping
FAR	False accept rate
FD	Fourier Descriptor
FOV	Field of view
FRR	False reject rate
GEI	Gait Energy Image
GFD	Generic Fourier Descriptor
HOG	Histogram of Oriented Gradient
ID	Identity
KNN	k-nearest neighbour
LBP	Local Binary Pattern
LDA	Linear Discernment Analysis
Lower H	Lower Half
LSTM	Long short-term memory
mAP	Mean average precision
mmW	Millimetre Wave Imaging
MHD	Modified Hausdorff Distance
MDTS-DTW	Multi-Dimensional Time Shift Dynamic Time Wrapping

MPFT	Modified polar Fourier transform
Person Re-id	Person Re-identification
RCD	Relative Distance Comparison model
RCP	Row and Column Profile
ROC	Receiver operating characteristic
SC	Shape Context Shape Descriptor
SVM	Support Vector Machine
Upper H	Upper Half
Upper Q	Upper Quarter
XQAD	Cross-view Quadratic Discriminant Analysis

List of Publication

- Rawabi Alsedais, Richard Guest, “Person Re-identification from CCTV silhouettes using generic fourier descriptor,” in *International Carnahan Conference on Security Technology (ICCST)*, 2017, pp. 1-6.

Chapter 1

Introduction

1.1 Intelligent Surveillance and Computer Vision

Monitoring cameras, such as closed-circuit televisions (CCTVs), are the common modalities for surveillance systems. Vast numbers of CCTVs have been installed to monitor critical public areas, such as airports, shopping malls, banking facilities, and college campuses. The CCTVs mainly monitor environments, people, events, and activities. These cameras provide an enormous amount of video footage that is conventionally monitored by enforcement officers. As human monitoring is prone to error due to factors such as fatigue, automated analysis of this video footage might improve the accuracy of surveillance efforts [1].

The field of computer vision through machine learning and artificial intelligence has recently inspired investments in automating manual surveillance systems to develop intelligent surveillance systems. *People analysis* is the umbrella term that involves research on various topics related to security and intelligent surveillance. This includes *human detection*, which is the process of finding the smallest bounding box that encloses each human in an image or video sequence. The detection process received a comprehensive exploration. In [2], multiple instances detection was examined, where in [3], environmental conditions of the surveillance were added to the human detection process. Various surveys have been conducted on this topic, such as [4] and [5]. Another currently hot research topic related to security and intelligent surveillance is *human tracking*, which involves tracking a person's movements from one frame to another to predict subsequent steps using the consistent temporal information [6], [7], [8]. *Activity recognition* is another topic gaining an impressive amount of attention in intelligent surveillance research [9], [10]. The aim of this field is to automatically analyse and classify human behaviour from one frame to another. A recent review on behaviour analysis can be found in [11].

The most recent area of computer vision and intelligent surveillance research to emerge is person re-identification (re-id). Person re-id can be defined as the process of re-identifying people observed in a camera's field of view (FOV) by matching them with people's previously observed identities [1]. When the newly observed person does not exist in the enrolled identities, this person's identity is automatically enrolled in the collection of previously observed people's identities. Therefore, previously observed people's identities comprise an evolving dataset into which new identities are added every time the system observes a new person.

In general, any system is able to recognise a person as long as the person stays in the same camera FOV and when the person's position, illumination, and background conditions are known to the system. However, in typical person re-id processes, the system is designed for a non-overlapping camera network; thus, a number of issues may arise, mainly concerning how the system identifies whether a subject in the current camera is the same subject seen previously in another camera of the same network. Therefore, the process of continually identifying people whose images were observed over a span of time and locations is called person re-id. Automatic person re-id for public surveillance systems poses many problems [12], as illustrated below.

- *Subject Detection*. A typical automated person re-id system requires a human detection application. The detection process is the first step of the person re-id; it can be defined as detecting people instances in a digital image (single shot) or video stream (multiple shots). This process is, in fact, an extremely sensitive part of any person re-id, as the accuracy of detecting a subject directly influences the accuracy of identifying and re-identifying that subject.

- *Subject Appearance*. The subject's appearance, such as colour, texture, shape, clothes, and pose, is a common way to build his or her unique signature, as it does not require any interaction with the subject. This signature is used as an identifier for this subject in the process of person re-id. However, human pose, and different camera characterisations in the same camera network may negatively affect the performance of these features.

- *Environmental Conditions*. Uncontrollable environmental conditions, such as subject occlusion, scene illumination, camera resolution, and viewing angle, are other challenges that negatively affect the stability of a subject's signature.

- *Distance Metrics*. In the person re-id process, the subjects' signatures are compared in order to be re-identified. Distance metrics are the responsible algorithms that compare signatures and assign subjects' identities to them. The typical scene of a person re-id system is an evolving scene with new subjects who are monitored in a camera's network. Simultaneously matching a large number of subjects with their identities from a camera's network is a significant issue, as the signature is built under different circumstances (i.e., locations and time). In addition, the increasing number of subjects leads to an increase in the signature similarity between subjects, making the distance metrics work more difficult.

It has only been a decade since person re-id application research started to become more prevalent. Considering the factors involved, as described, person re-id remains an unresolved challenge in surveillance system and computer vision research and design.

The remainder of this chapter presents the motivation for this research and is organised as follows: Section 1.1 presents a brief introduction to intelligent surveillance and computer vision; the motivation for and objectives of this research are listed in Section 1.2 and 1.3, respectively. Section 1.4 outlines the study's contributions, and Section 1.5 describes the overall thesis structure.

1.2 Research Motivation

The increasing demand for public security, including forensic security, has resulted in a considerable growth in the presence of surveillance camera networks (CCTVs), including critical infrastructures, in public areas. Analysing the data gathered from CCTV footage is an essential part of evaluating people's behaviour and activities captured on these networks. This analysis

enables multiple responses to suspicious events, such as real time alarms, incident retrieval, and increased security team awareness.
Significant improvements in the computer vision and machine learning fields have advanced traditional surveillance camera network systems (i.e., manually monitored by people) towards the development of intelligent surveillance systems that have enhanced the monitoring process through automated person detection, person tracking, activity recognition, and person re-identification. Consequently, this shifts the focus from post-reaction and incident retrieval to event prevention.

Person re-identification (re-id) techniques can substantially improve intelligent surveillance. In the literature, many techniques have been proposed to examine various appearance-based features, such as colour, texture, and clothes. However, the use of the body and body parts shape as identifiers for person re-id remains subject to examination, prompting this research study.

The system proposed in this study employs the extracted features of the body and body parts, utilising one of the well-known shape descriptors to be further processed and used as a unique signature for each subject in a scene. In order to assess the feasibility of the body shape descriptor, this signature is employed in two different systems. The first is the image-based system, which classifies signatures on a frame-by-frame basis. This system examines the discrimination level of the static variation in the body shape descriptor. The second system is the video-based system, which exploits the signatures of the entire sequence to examine the discrimination level of dynamic variations of the body shape descriptor.

1.3 Research Objectives

The objectives of this study are as follows:

1. Determine if the body shape descriptor contains discriminative information for person re-id.

2. Identify any factors that negatively affect the discrimination levels of the person re-id based on the body shape descriptor and determine to what extent these factors are influencing the performance.

3. Investigate the discrimination levels of different parts of the body by applying a system of body segmentation, by examining whether the shape descriptors of different parts of the body could yield different levels of discriminations.

4. Investigate the possibility of integrating (i.e., fusing) the results of the frames collection of one sequence for performance enhancement.

5. Determine the dynamic feature that can be exploited in the shape descriptors of the frames of one sequence to classify subjects based on.

6. Determine whether the dynamic feature can be exploited via the shape descriptor of the frame collection of a sequence to classify subjects.

7. Compare the performance of the developed system with published works on the same dataset.

1.4 Contributions

This thesis makes five main contributions to the shape-based person re-id, which are summarised as follows.

First, a body segmentation mechanism is proposed to investigate the discrimination of the body and body parts shape descriptors for person re-id. This segmentation is designed based on multiple anthropometry studies.

Second, shape descriptors of the body and proposed body segments are used as identifiers for person re-id. A Generic Fourier Descriptor (GFD) 4.2 is used to generate an individual shape descriptor for each proposed body segment. These descriptors are exploited for the image-based system implementation. The re-identification process is implemented on a frame-by-frame basis in this system using Linear Discriminant Analysis (LDA) 4.2.3. This, in fact, assesses the feasibility of re-identifying a subject based on the subject's shape static feature.

Third, re-identifying subjects using the dynamic features of the shape descriptors of the subject's body and body segments is explored. The GFD shape descriptors of the body and body parts are utilised in this experiment. The multidimensional Dynamic Time Wrapping (DTW) 5.2 algorithm is used on a sequence-by-sequence basis to re-identify subjects based on their dynamic variation.

Fourth, comprehensive system outcomes of the image-based and video-based systems are analysed. The analysis compares the performance of both systems for each segment individually, revealing several significant findings.

Fifth, the rank list fusion method 6.3 is proposed for performance enhancement for several aspects of this research. The fusion method combines the image-based generated rank lists so that the lists generated by all frames in each sequence are replaced by one rank list. The main purpose of the proposed rank list is to count the indices of each identity in the initial rank lists. The indices of each identity from each initial rank list are added together, and then the identity with the least total indices is placed in the frontal location of the new fused rank list. This method is implemented in four appearance related approaches including wearing clothes and bag.

1.5 Structure of the Thesis

This thesis is organised into seven chapters, which can be briefly summarised as follows:

Chapter 1 introduces the topic of person re-identification and clarifies its place within biometric systems. This chapter also illustrates the motivation for and contributions of exploring the use of the body and body segments shape descriptors for the purpose of person re-identification. Finally, it presents the structure of the thesis.

Chapter 2 reviews the soft biometrics extraction methods and the method for using these features for person re-identification. It also describes revisions of the available methods for body segmentations and their advantages and disadvantages.

Chapter 3 introduces the proposed shape-based person re-id system's framework and discusses the stages involved in this system. This chapter also includes a review of the publicly available person re-identification datasets and justifies the reason behind using CASIA Dataset B 3.3. The proposed body segmentation method is presented in detail in this chapter, including the arithmetic operations that have been applied on the body silhouette in order to generate the proposed body segments. The segments' length parameters are justified based on four anthropometrical studies in this chapter as well.

Chapter 4 introduces the methodology for extracting the shape descriptor from the body and body segments shapes and for using them for re-identification, using a classifier that sorts them on a frame-by-frame basis. In addition, this chapter presents an examination of the inter-view scenario, where the subject's identified shooting angle differs from the re-identified viewing angle. Moreover, the cross-scenario approach is inspected, where the subject's appearance changes with the addition of wearing a coat or carrying a bag. The classification results of all mentioned scenarios are discussed in this chapter.

Chapter 5 exploits the shape descriptor features of the body and body segments to examine person re-id based on the subject's dynamic variation. This is applied using a time series analysis algorithm, and the re-id process is applied on a sequence-by-sequence basis. The results of this implementation are presented and discussed in this chapter. The inter-view and cross-scenario approaches are examined using the video-based system in this chapter, also.

Chapter 6 provides a comparison and discussion on the outputs of the image-based and video-based systems outcomes. It also describes the proposed rank list fusion mechanism that exploits the multiple rank lists for each sequence to generate improved performance through one fused rank list for each sequence. The same mechanism is applied on the inter-view and cross-scenario approaches in pursuit of performance enhancements.

Chapter 7 highlights the contributions that are added to the field of person re-id research by this study. It also lists several limitations faced in this experimental work. Finally, it discusses further work needed to enhance the performance of shape-based person re-id.

Chapter 2

Literature Review

2.1 Introduction

The main goal of this research is to investigate the power of using the shape of different parts (or segments) of the body to re-identify (re-id) human subjects in public areas; this will be assessed through image-based system (i.e., static features) and video-based system (i.e., dynamic features) contained within each segment shape descriptor. Drawing on the literature, this chapter explores three main areas undertaken in this research, namely *Biometric Systems, Person re-id and Body Segmentation.*

The structure of this chapter is as follows: Section 2.2 provides a brief overview of biometric systems and the general performance evaluation. Section 2.3 focuses on what is person re-id and how fits the overarching concept of biometric modalities and use cases. Similar fields that overlap the concepts of person re-id are also illustrated. Also, it presents a general person re-identification (re-id) framework to address specific scenario-based challenges, details a number of ways of representing and classifying subjects and outlines deep learning methods and common evaluation metrics for the field. Section 2.4 reviews the concept of *Body Segmentation* in recognition systems, elaborating different techniques of segmenting individuals. Section 2.5 concludes the literature review, states the opening research questions and recommends areas for further development.

2.2 Biometric Systems

A biometric system can be either a *verification* system or an *identification* system [13]. In a *verification* configuration, the subject claims a particular identity, and then the system compares the provided data with the stored template data of the claimed identity. This means that the verification process is a one-to-one mode, which tests the authenticity of the claimed identity. Biometric systems that operate by verification are common and examples include verifying the face [14], [15], fingerprints [16], [17] and iris [18], [19]; the first two modalities are widely used in the market.

For *identification*-based configurations, the system recognises a subject by selecting the best match between the provided data and the data of each enrolled subject in a dataset. Thus, the identification process is considered a one-to-many comparison process for identifying a subject, without the explicit claim to an identity. Identification biometric applications are mostly relevant to forensic security [20]. Building security and door access control systems can recognise subjects by their voices [21]. The face is another biometric that can be used in identification [22] for law enforcement and surveillance.

The performance of verification and identification systems is indicated using the receiver operating characteristic (ROC) curve. A ROC curve represents error rates, including false accept rates (FARs) and false rejection rates (FRRs). An FAR reflects a mistaken match of two biometric samples belonging to different persons; an FRR reflects the mistaken rejection of two biometrics samples belonging to the same person.

Person re-identification (re-id) is another recent topic in biometric systems, which is the main focus of our research. A detailed survey on person re-id is conducted in Section 2.3 of this chapter.

2.3 Person Re-id

Public areas (such as airports, train stations and shopping malls) has received increased attention within computer vision research with the aim of enhancing the security levels. Installing and utilising CCTV networks within non-overlapped field of view provides enhanced coverage. These cameras provide an enormous amount of video footage that conventionally is manually monitored by enforcement officers. As human monitoring is error prone through factors such as fatigue, automated analysis of these videos might improve or support the accuracy of surveillance.

The person re-id process, which is a relatively new area of research in computer vision and intelligent surveillance, is one method for automating CCTV image monitoring and analysis. As a biometrics system, person re-id can be defined as the process of re-identifying people observed in a camera's FOV by matching them with previously observed people's identities using their signatures. When the newly observed person does not exist in the enrolled identities (i.e., watch list), this person's identity is automatically enrolled in the previously observed people's list. Therefore, this list is an evolving dataset, as it adds new identities every time the system observes a new person. A person re-id application primarily involves two main procedures: first, the methodological representation of the person in a camera's FOV (i.e., constructing the signature); second, the evaluation of similarities between the person's digital representation and those of previously observed subjects in non-overlapping camera networks that have no common FOV.

When a re-identified (i.e., previously identified) subject disappears from a camera and re-enters the FOV of the same and/or another camera, the person re-id application should be able to determine that the subject had been observed previously and match the subject's identity with the existing identity utilising the subject's extracted features, or signatures. Therefore, the matching process, i.e., determining what previously observed identity is closest to the identity of the person being observed (who had already been seen), operates according to similarities in the features (or digital representations) that have been extracted from a single image (image-based approach) or multiple images (video-based approach).

To highlight, the main differences between identification and re-identification applications are a) the identification system dataset contains a finite set of subjects, such as fingerprints and DNA datasets, unless manually updated, while the person re-id system dataset continually evolves as newly observed subjects are automatically added; b) identification configuration is a one-to-many process, as previously explained, while person re-id is a many-to-many process that compares all observed subjects with enrolled identities in an attempt to find a match.

A close inspection of people analysis applications, especially person re-id, tracking people, and query-based image retrieval, may blur the boundaries between each. These concepts share several constraints and interfere with each other. **Tracking,** for instance, uses consistent temporal information, which means appearance continuity is required to follow a subject's movements from one frame to another, mainly to predict the subject's next moves [2], [6], [7], [8]. This prediction is primarily based on the temporal information [23] and assumes that the FOV overlaps across cameras. In contrast, the purpose of person re-id is to match subjects' identities, even if there are time delays and/or FOV changes [20].

Image retrieval also shares a similar range of processes with person re-id [23]; they both seek to identify possible instances of a particular person. In fact, there are some approaches that can be applied to both image retrieval and person re-id [24]. However, **image retrieval** mainly focuses on searching for a digital image in a dataset based on a query of another digital image or semantic query provided by an end user, while person re-id continuously compares all image streams with existing previously recognised identities. Therefore, in **image retrieval**, all possible queries are learnt during the training process, whereas in person re-id, the application should learn about the similarity metrics of any given pair of image instances [12].

2.3.1 Challenges of Person Re-identification

In general, a system is able to recognise a person as long as they stay in the same camera FOV and when the subject position, illumination and background conditions are known to the system. However, in situations where the system is designed for a non-overlapping camera network, a number of issues may arise, mainly concerning how the system identifies whether a subject in the current camera is the same subject seen previously in another camera of the same network. The process of continually identifying people who are spread over time and location is called person re-id. Therefore, automatic person re-id for public surveillance systems poses many problems [12], as illustrated below.

Re-id by humans is a straightforward action that is easily carried out on a daily basis. Humans are able to easily extract the features based on a person's appearance (e.g. face, clothing, hair, voice and gait) and later re-identify them based on their descriptions. Re-id systems, however, require unique features extracted from the subjects in the scene, and an accurate corresponding signature matching. Therefore the process of automating person re-id poses many issues, which are mainly related to either the descriptors and/or the similarity metrics [12], [25].

Any person re-id system can be divided into three main stages: (a) detecting the subject and segmenting the body into the identified sections [25], (b) building unique signature on the selected part(s) [26] and (c) learning a similarity metric which minimises intra-class and maximises inter-class differences.

Figure 2.1 illustrates a general framework of an automated person re-id system. Various tasks must be completed during each phase, which makes automating each stage of the process more challenging.

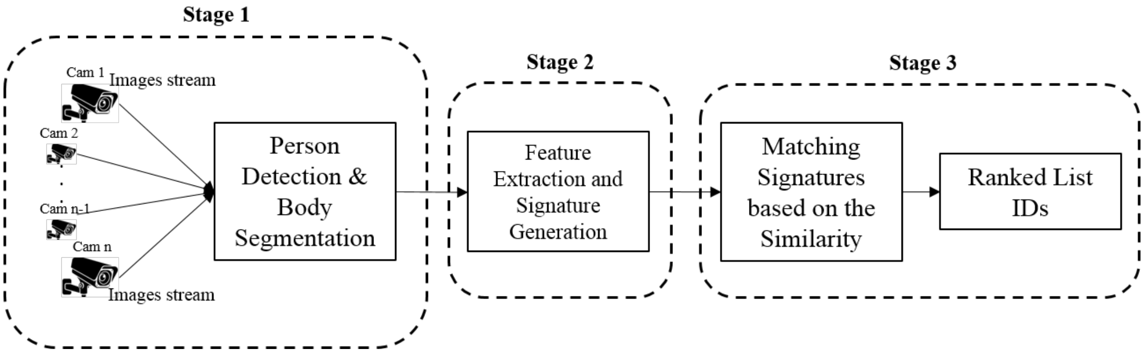

Stage 1
Cam 1
Images stream
Cam 2
Cam n-1
Cam n
Images stream
Person Detection & Body Segmentation
Stage 2
Feature Extraction and Signature Generation
Stage 3
Matching Signatures based on the Similarity
Ranked List IDs

Figure 2.1: Person re-id general framework

Stage 1. In the first stage of a typical automated person re-id application, the subject must be detected, and then segmented into the concerned parts. The purpose of the detection process is to identify people instances in a digital image (single shot) or video stream (multiple shots). Person detection is a research field with unique issues and challenges; thus, it has been studied for decades [27], [4], [28], [2]. Automatic subject detection is a fundamental process to introduce an applicable re-id system to be applied on the field. However, the automatic detection process is a substantial field on its own in computer vision that could affect person re-id performance. Therefore, the research in person re-id tends to avoid the involvement of automatic subjects detection as it interferes with person re-id negatively. Instead, the studies in person re-id tend to manually detect the subjects in the image frames. The other process conducted during this stage is body segmentation, which is performed only if a part of the body will be included in the representation and feature extraction. Body segmentation is discussed in detail in section 2.4 (Body Segmentation) of this chapter.

Stage 2. The second stage of the re-id system is to extract features and to build a unique signature for each subject in the scene. The appearance based features of a person is the most obvious features that can be extracted from a video sequence, for instance, colour, texture and shape. However, the scene illumination, human pose and different camera characterisations in the same camera network may negatively affect these features performance. Clothing is another widely known appearance based descriptor, which can be a distinctive feature of short-period systems; thus, it may not be very suitable for long-period systems, which compare correspondences that are captured days or months apart. Besides, it is possible that people are dressed alike, which may lead to assigning the same features to different people. Additionally, the uncontrollable environmental conditions, such as subject occlusion, camera resolution, and viewing angle, are other challenges that need to be taken into account. Generally speaking, for any feature (or digital representation), factors exist that lead to maximising intra-class differences, which negatively affects the matching process between one subject's instances.

Stage 3. In the final phase of person re-id, subjects' digital representations or descriptors are compared to find the closest match, with a view to generating the identities ranked list. Even if the features are effectively extracted, simultaneously matching a large number of subjects with their identities from a camera network is a significant task, as the features are extracted under different circumstances, e.g., locations and time. Also, as the number of subjects in the scene increases, the features specificity decreases, leading to an increased possibility of false matches.

It has only been a decade since attention to person re-id applications has been increasing, and, recalling the above-mentioned issues, person re-id remains an unresolved challenge in surveillance systems and computer vision. The Sections bellow review the state-of-the-art of representation approaches and similarity metric learning methods.

2.3.2 Evaluation Metrics

Before reviewing the feature extraction methods (i.e. representation) used for person re-id applications in this section, two evaluation metrics are discussed, as the evaluation metrics are used with each representation method to illustrate its performance accuracy.

The most common evaluation metric used with re-id systems to evaluate and compare performances is the cumulative matching characteristics (CMC) curve [29]. CMC curves accumulate the number of the true re-id for each rank and show them on order. The CMC calculation of the true re-identifications in rank i is:

$$CMC(i) = \sum_{r=1}^{i} tq(r) \tag{2.1}$$

where tq is the true re-identified queries at rank r. For simplicity, a brief example is presented next: given n identities (classes) and one frame including unknown identity subject (i.e. test frame) that belongs to the n identities. This frame is compared with each class and produce one similarity score, where it estimates the similarity between the unknown subject and the tested class. The comparison of the tested frame with all the classes generates a list of n scores, one score for each class. Then the list is arranged in descending order, where the larger score means the more similarity between the tested frame and the tested class and vice versa. This means that the score of the most similar identity is located at the beginning of list which called first rank. Also, the score of the least similar identity is located at the end of the list which called nth rank. Therefore, this list called rank list. Then, the scores in the list are replaced with their corresponding identities. Then the identity in each rank is compared with the ground truth identity of the subject in the test frame. If they match, the true match counter of the current rank increases by one. If they do not match, then the true match counter of the current rank remains the same. The equation (2.1) applies the same concept of the given example but on all the frames in the test set instead of one frame.

One main advantage of CMC curves is that it can show the number of true re-identifications across all ranks. This allows the CMC curve to indicate the steep of the curve, illustrating the quality of the performance, where a steeper curve shows a better performance. Another evaluation metric that has recently been used is the mean average precision (mAP) evaluation [30]. This evaluation metric is used with CMC in cases where the ground truth is more than one sample. In these cases, mAP evaluates and considers the distances between the true matches where CMC fails to do so.

2.3.3 Feature Extraction

Appearance-based approaches are widely used in the literature to establish a similarity between correspondences [12]. These approaches extract features from the subjects' appearances, such as clothing colour, texture and type. In the next subsections, number of appearance-based model features are reviewed.

2.3.3.1 Colour and Texture

Colour spaces represent the image colour as a numerical value of certain bases. RGB [31] represents the red, green, and blue. Lab [32] expresses lightness, while a and b denote green–red and blue–yellow colour components. YCbCr [33] is another colour space, where Y stands for the luminance, Cb for the blue difference, and Cr for the red difference. HSV [34] represents the hue, saturation, and lightness of the RGB. These well-known colour spaces have been widely used in person re-id.

In [35], an experiment was conducted on VIPeR dataset to illustrate the discrimination level of these colour spaces. The VIPeR dataset includes two cameras, each of which provides one image for each subject, with 632 total identities. In order to evaluate the performance of different colour spaces, a random forest evaluation method was built for each pair of images. This method evaluates the similarity of each pair images using the features of 6 colour spaces, namely RGB, normalised RGB, HSV, YCbCr, CIE XYZ and CIE Lab. The similarity function was then learnt for those images in order to match them.

Table 2.1 shows the matching rate at the 1st, 5th, 10th and 30th ranks. From the table, it can be seen that HSV colour space outperforms other colours. However, the accuracy rate of using a colour space on its own for re-id application, in general, is considered low, mainly because colour spaces between different subjects tend to resemble one another and are affected by illumination. Therefore, colour spaces, as a representation, tend to be combined with other kinds of features.

Table 2.1: Matching rate (in percentage) of four common colour spaces at different ranks [35]

Colour spaces	r = 1	r = 5	r = 10	r = 20	r = 30
RGB	2.50	10.60	16.39	23.67	28.54
HSV	12.63	26.11	35.10	46.11	54.81
YCbCr	10.60	23.67	32.69	43.86	52.72
Lab	11.08	25.73	31.87	39.91	46.87

Texture information is another common feature that is combined with the colour features. For example, in [36], the HSV colour of the images and a Histogram of Oriented Gradient (HOG) [37] of the texture were extracted as combined descriptor in the training and testing sets of images. This means that the HSV feature combined with the HOG feature to create one descriptor for each image in the training and test sets. The similarity between the probe images descriptors and gallery image descriptors are measured using Cosine similarity. The main reason for using the cosine similarity is the efficiency of computing the features of high dimensions within a large dataset. Figure 2.2 shows the original images and their HSV correspondences.

The experiment was conducted on three datasets, namely, VIPeR, ETHZ(SEQ2), and CAVIAR4REID. The ETHZ dataset images were collected from a moving camera with a small viewpoint variance. The illumination and scale variance and the occlusion level are considered high compared with other person re-id datasets. The CAVIAR4REID dataset images were collected from two surveillance cameras at a shopping mall with overlapping FOVs. The dataset includes 72 identities, 50 appearing in both cameras and the rest appearing in one of the two cameras. The experimental results demonstrated that this method outperformed the PRDC [38], ICT [39], and SDALF [40] methods.

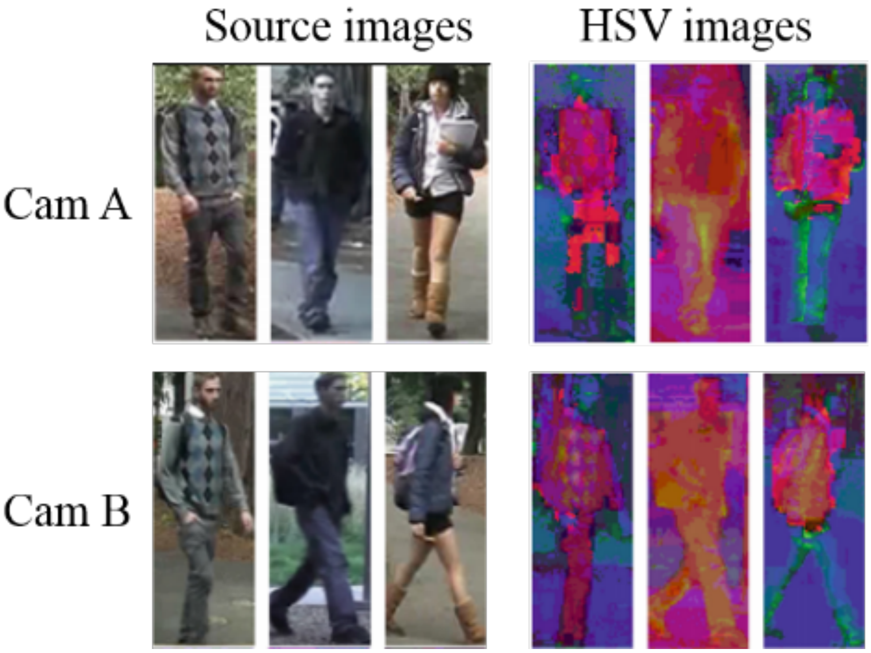


Figure 2.2: Source images from VIPeR dataset and the HSV colour space of the same images [36]

In [41], HSV was also used with Lab colour space and Local Binary Patterns (LBP) as a texture descriptor. The experiment was conducted in VIPeR, PRID 2011 and ETHZ datasets. The PRID 2011 dataset images were collected from two cameras, A and B; camera A included 386 trajectories, and camera B included 749 trajectories. Two hundred subjects appeared in both cameras.

Every image in these datasets presents a tight bounding box that only contains one subject. For this study, the image was sampled into 8x16 rectangular regions with a size of 4x8 pixels. For each rectangular patch, the average value of the colour channel was computed and then discretised to the range of 0 to 40. The texture information of each patch was also extracted using LBP. The colour and texture values were concatenated as one feature vector. All the vectors from different patches were then linked to generate one vector that represented that image. Figure 2.3 shows the global feature vector of one image.

In the classification stage, it was assumed that the probe image view was different from the gallery image views. The similarity between the probe images and gallery images was then computed using a proposed learning metric, which returned a gallery image with the smallest distance as a potential correspondence to the prop image.

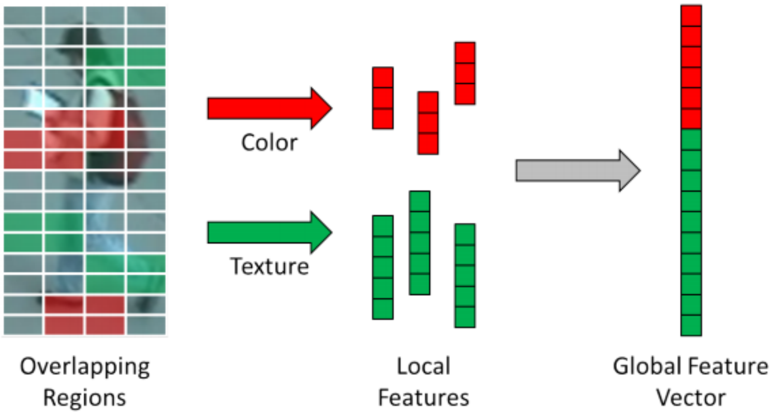


Figure 2.3: The image was sampled into rectangles patches. with a global image feature vector consisting of fused features (HSV, Lab, LBP) of each patch in the image [41]

As RGB colour space is sensitive to lighting, in [42], using RGB infrared (RGB-IR) imaging in re-id was addressed by matching the RGB image with RGB-IR images. A new re-id dataset, called SYSU-MM01, was created, which included RGB images (the original image) as well as RGB-IR images from six cameras with 491 IDs. In total, 296 identities were selected for training, 99 for validation and 96 for testing. RGB images from cameras one, two and four were given as a gallery set and RGB-IR images from the cameras three and six were for probe set. Then, the similarity between each RGB-IR probe image and the RGB gallery images was computed. Next, a ranking list for each prop image was generated. The experiment showed that, despite comparable results, matching RGB images with RGB-IR images was a challenging process. Figure 2.4 shows images of RGB and IR taken by day and night.

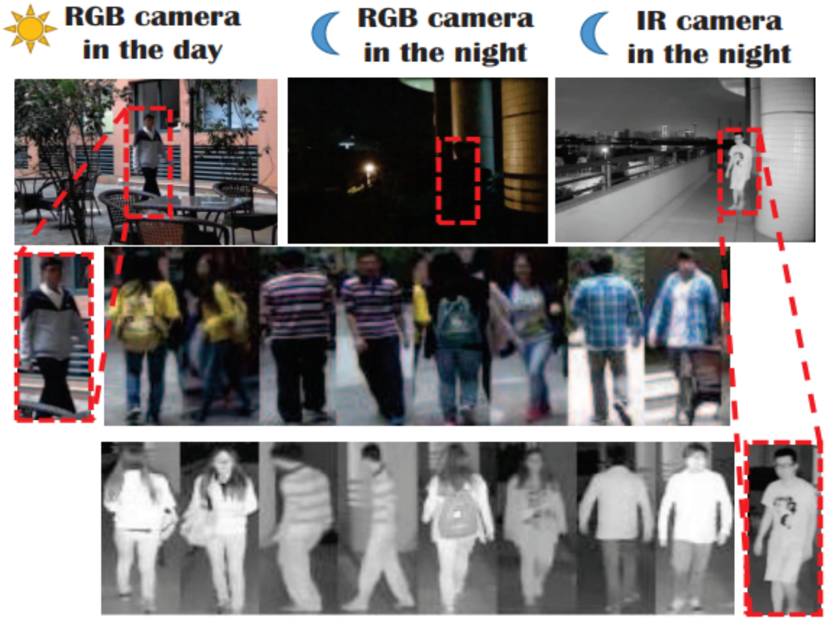


Figure 2.4: The RGB images by day and night and the IR images by night [42]

Finally, the main advantage of using colour and texture as descriptors or distinctive representations for person re-id lies in the convenience of extracting them in terms of the detection or the computational cost. The main drawback to these descriptors is the negative influence illumination has on their performance. Moreover, because people tend to dress differently on different days, clothes colour and texture features are more suitable for short-term person re-id applications. Combining these features with illumination invariant features and features that do not change over extended periods of time, such as body shape features, may be one way to address this issue.

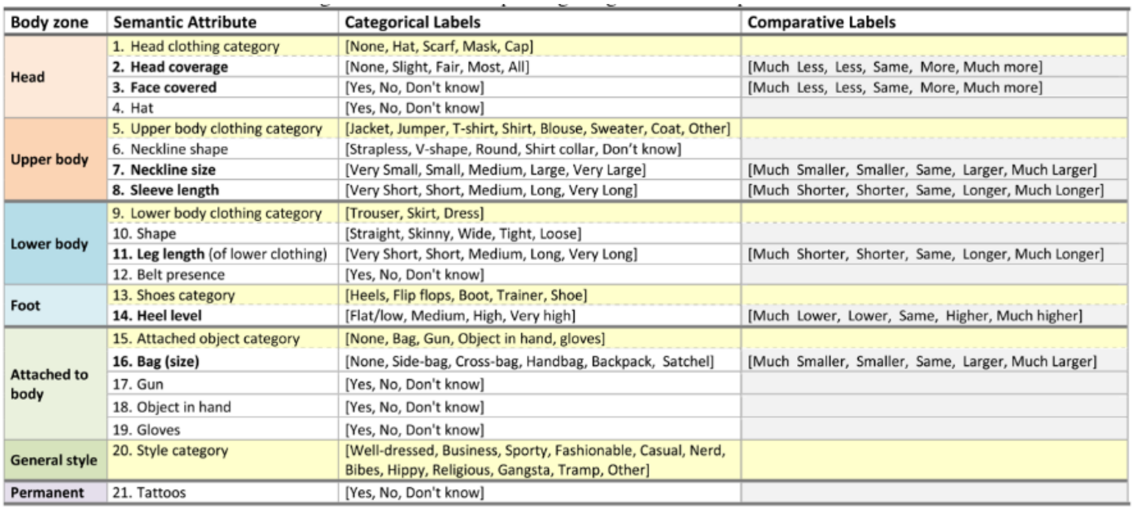
2.3.3.2 Clothes

Compared with colour spaces and texture information, clothing types as an identifier have not been deeply explored [43]. In [44], human's soft clothing attributes were analysed to explore their enhancement in subject retrieval. For human identification, in [45], the body was divided into seven different zones; each zone was then assigned different semantic attributes, categorical labels and comparative labels (see Table 2.2).

The Soton Gait Database [46] was used in this study by developing a web-based system that manually collect clothing attributes and comparisons of the subjects appearing in the Soton Gait

Dataset. The attribute collection process was completed by asking a number of users to describe a set of subjects from the dataset by selecting one label for each attribute. Using clothing comparison for re-id was not sufficient. However, subjects were listed according to one attribute; then, each subject was described using these ordered lists. This was accomplished by using soft-margin Ranking SVM method [47]. Clothing attributes were then augmented with body soft biometrics, which were explored in [48] that included age, gender, ethnicity and skin colour attributes. The performance of soft biometrics and clothing attributes augmentation was evaluated using CMC curve, which illustrated a 75% accuracy rate at the first rank and a 100% accuracy rate at rank 29.

Table 2.3: Body zones and their semantic attributes, categorical and comparative annotations [45]

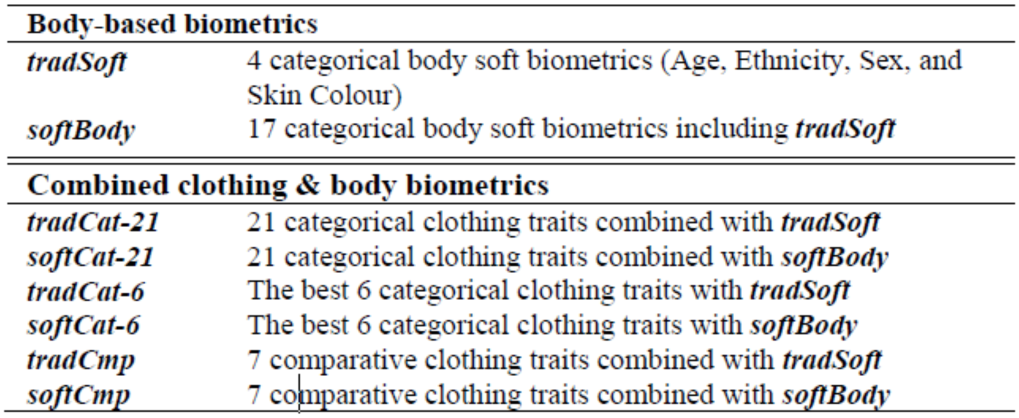


Body zone	Semantic Attribute	Categorical Labels	Comparative Labels
Head	1. Head clothing category	[None, Hat, Scarf, Mask, Cap]	
	2. Head coverage	[None, Slight, Fair, Most, All]	[Much Less, Less, Same, More, Much more]
	3. Face covered	[Yes, No, Don't know]	[Much Less, Less, Same, More, Much more]
	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	[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]	

Clothes have also been used to address the subject viewpoint variation in [49] to aid re-id applications. In the same context as in the previous study, an observed subject from multiple views is described by a verbal query. This query is used as a description to the probe image, which is then used to compare it with the gallery image descriptions. It should be noted that the probe image viewpoint is not included in the gallery image viewpoints. The query is constructed based on a group of clothes related traits integrated with other traits as shown in Table 2.4. The ***tradSoft*** in this table are the traditional soft biometrics including age, ethnicity, sex ans skin colour. The ***softBody*** traits are 13 body soft biometrics shown in Table 2.5 and the four *tradSoft*. The **tradCat-21** traits are the categorical labels of the 21 traits in Table 2.3 combined with *tradSoft*. The **softCat-21** traits are the categorical labels of the 21 traits in Table 2.3 combined with *softBody*. **tradCmp** and **softCmp** are the seven comparative labels in Table 2.3 combined with *tradSoft and softBody respectively.*

Figure 2.5 shows one query image example and its correct retrieval image at the first rank and another query image with its correct retrieval image at the 7th rank. The CMC curve shows that the categorical clothes group (SoftCat-6), including Head coverage, Sleeve length, Leg length, Neckline size and Heel Level, outperforms the other comparatively and categorically based groups. SoftCat-6 achieved 0.94 at the first rank, when it was combined with soft body biometrics, consisting of age, gender, ethnicity and skin colour attributes.

Table 2.4: Different groups of clothes related traits [49]



Body-based biometrics	
tradSoft	4 categorical body soft biometrics (Age, Ethnicity, Sex, and Skin Colour)
softBody	17 categorical body soft biometrics including ***tradSoft***
Combined clothing & 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***

In the same experimental context, the clothing attributes were fused with face and body features for person recognition in [50]. Clothes, face and body traits (i.e. age, gender, ethnicity and skin colour attributes) were collected using the web-based system by asking the participants to compare and categorise the attributes the subjects appearing the Soton Gait Dataset. In the implementation stage, the Euclidian distance was used between the prop image collected attributes and all gallery images collected attributes. For recognition, the image with the smallest distance was considered the correspondence identity to the probe image. For verification, the target gallery image was required to meet the predefined threshold or was rejected otherwise. For evaluation, Figure 2.6 shows the individual performance of the categorised and comparative body, face and clothes attributes. The three different modalities are then fused under two scenarios, shown in Figure 2.7.

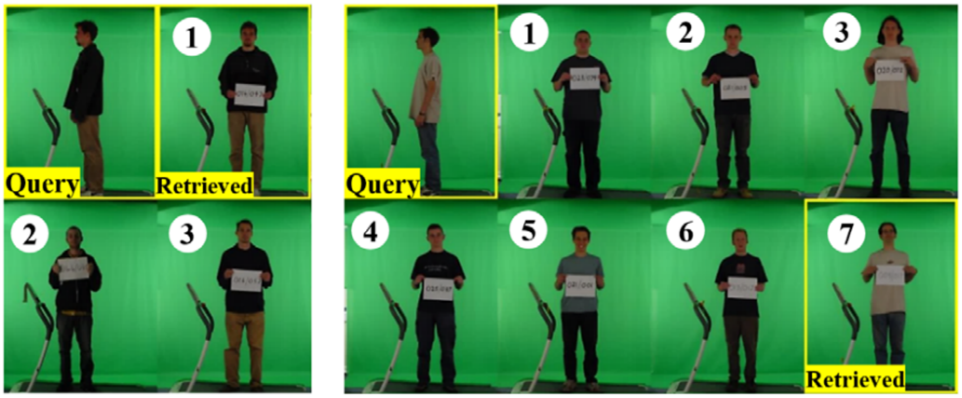


Figure 2.5: Two different queries and their correct retrieval ranked images [49]

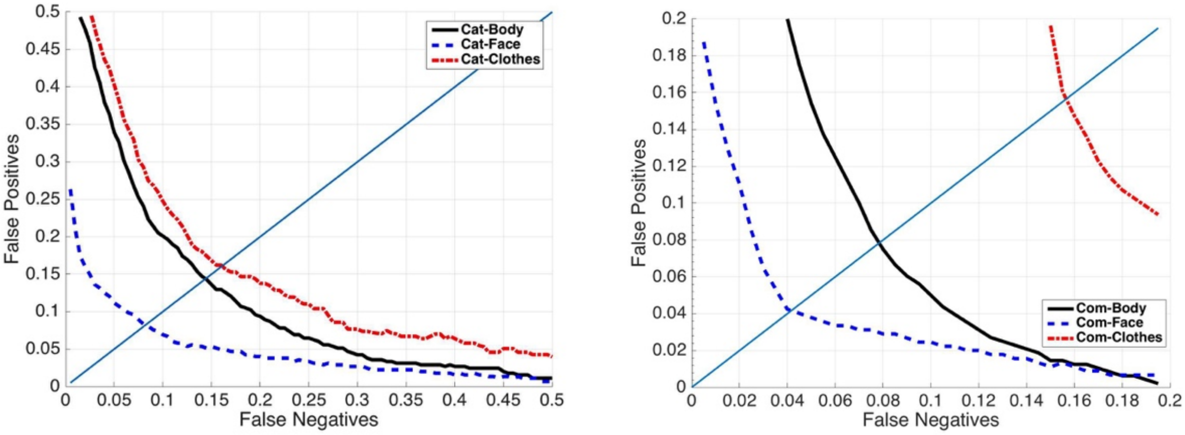


Figure 2.6: The separate performance of three different modalities, where (a) is the ategorical Body, Caterogical Face and Categorical clotheing attributes and (b) is the Comparative Body, Comparative Face and Comparative Clothes attributes [41]. Categorical and comparative labels can be found in

Table 2.3.

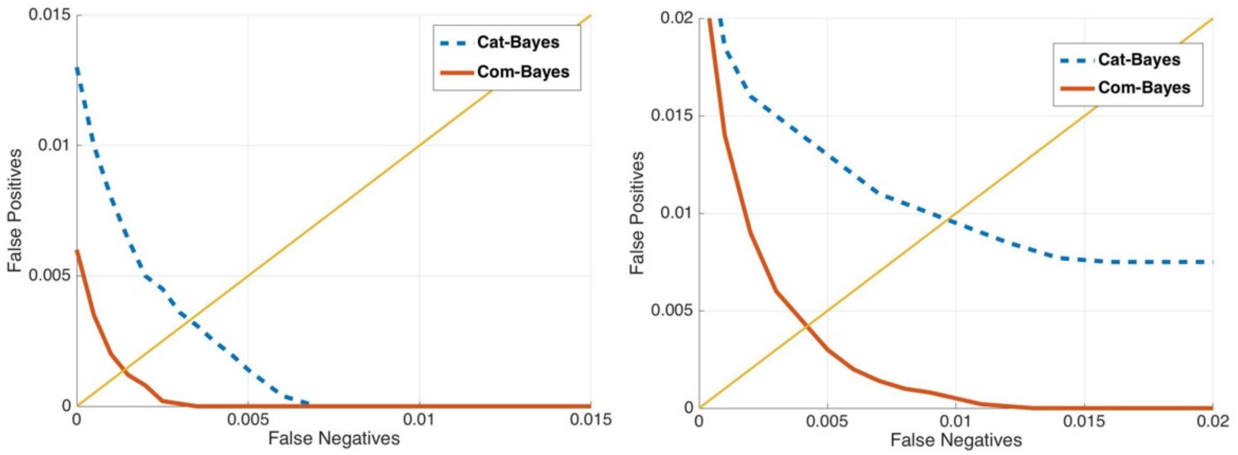


Figure 2.7: The performance of fusing the categorical and comparative attributes listed in Table 2.3 with the face and body traits using the probability density by Bayes theorem : (a) Fusion of the body, face and clothes (b) Fusion of body and clothes only without the face [50]

In general, these studies showed that clothing attributes can be used in association with other soft biometrics for human identification performance enhancement. Clothes attributes, however, have been interpreted based on users' manual observations, which means that the labels have been compared and categorised by users. Therefore, clothes attributes are not yet suitable for implementation for person re-id in public areas, as person re-id requires an automated representation of all subjects in the camera's FOV. Consequently, the clothes attribute needs to be recognised, extracted, and analysed automatically so that it can be integrated with other appearance-based features for person re-id.

2.3.3.3 Body shape

Describing the shape of the body is another soft biometric, which has been used as a digital representation to identify subjects of interest. One way of arithmetically describing a shape is to use one of the wide range techniques of shape descriptor. Shape descriptors, in general, have been reviewed in [51] and [52]. However, no evidence has been found in the literature of any technique other than the *Shape Context (SC)* shape descriptor and *Fourier descriptor (FD)* (of shape descriptors) being used for person recognition purposes. The studies examining these methods are discussed in this subsection.

For person recognition, there are only a few studies that used shape descriptors to represent the human body shape. For example, in [53], a person was recognised through Millimetre Wave Imaging (mmW), using a combination of body shape description and texture information. The recognition system is illustrated in Figure 2.8. Texture representation was implemented using Local Binary Pattern (LBP) and Histogram of Oriented Gradients (HOG). Shape was represented across a number of processes: (a) Contour Coordinate (CC) is the baseline feature that gives the coordinates of the pixels, located at the edges of the silhouette. It can be defined as: $CC(n) = (x_n, y_n), n = 1, \dots, n_{cc} - 1$, where n_{cc} is the number of pixels; (b) Row and Column Profile (RCP) are computed as the number of pixels in the row and the column, which belong to the silhouette; (c) *Shape Context (SC)* shape descriptor describes each point at the edge by the angle in accordance with the distance to all the other points. The number of radial bins and theta bins are two parameters that should be pre-determined. Shape context was introduced in [54].

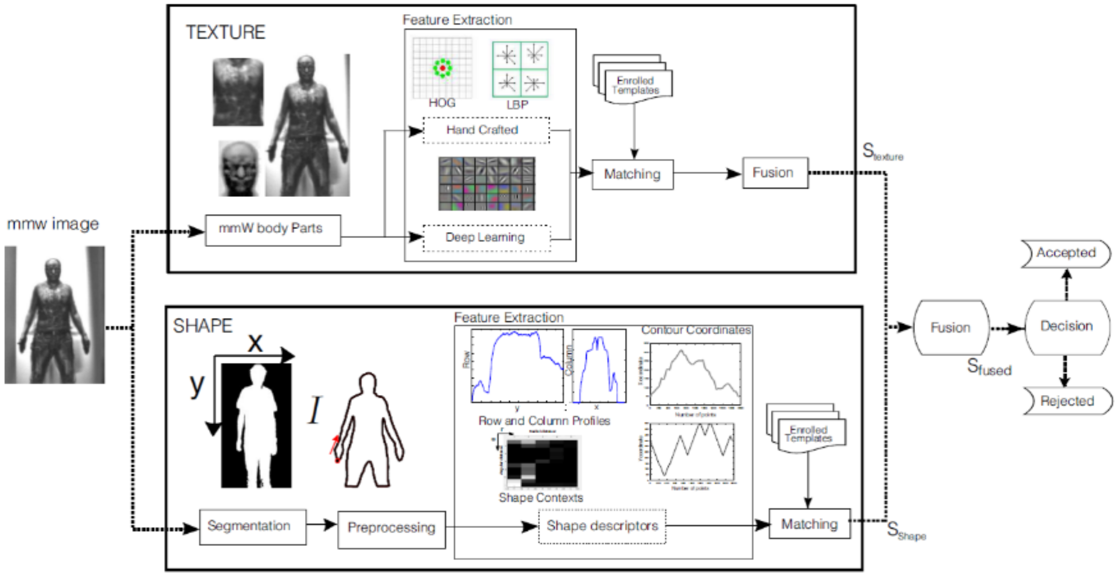


Figure 2.8: The system in [53], showing the fusion of the texture and shape descriptions

For matching, Dynamic Time Wrapping (DTW) and Modified Hausdorff Distance (MHD) were used, as they are two matchers that measure the similarity between the two sequences. The results showed that the texture-based feature always outperformed the shape-based feature in this particular study. However, after configuring different fusion scenarios, the best results were either Contour Coordinate (i.e. the baseline feature that gives the coordinates of the pixels, located at the edges of the silhouette.) with MHD or Row and Column Profile (i.e. computed as the number of pixels in the row and the column, which belong to the silhouette) with Dynamic Time Wrapping (DTW), reaching EER = 1.50%. *Shape context* (SC) descriptor was also used in [55], using Canny edge technique to detect the edges.

Fourier Descriptor (FD) [56] is another shape descriptor that has been used as a shape representation for person recognition. FD is a contour-based descriptor, which is translation, scale and rotation invariant. In [57], FD was one of the features that was used to explore body shape from millimetre waves (mmW) attained images for person recognition. For each pixel located at the edge of the subject body shape, the coordinate (x_n, y_n) was attained. The FD $f(l)$ was obtained thus:

$$f(l) = \sum_{n=0}^{n_{cc}-1} u_n \exp\left(-j\frac{2\pi}{n_{cc}}\, ln)\right), l = 0, 1, \dots, n_{cc} - 1 \qquad (2.2)$$

where $u_n = x_n + jy_n$ and n_{cc} is the number of pixels located at the shape contour. DTW and MHD were also used as distance-based matching techniques. Although the results showed that FD performed better with DTW than MHD, the performance of the FD with SVM classifier outperformed DTW and MHD. The results also indicated that the accuracy of FD was the lowest, compared with the other features, such as contour coordinate CC, row and column profile RCP and shape context.

The experiment was conducted based on two scenarios: the first scenario involved either a frontal head pose or a lateral head pose, and the second, the cross-pose, included two poses for the purpose of examining the features against subject pose variations. For frontal side body orientation, the FD shape descriptor obtained at the first rank 48.50%, 36.50% and 54.00% for DTW, MHD and SVM, respectively. For cross-pose orientation, the results were 54.00%, 34.00% and 52.50% for DTW, MHD and SVM, respectively.

The accuracy of the body shape description was fused with that of other features for person identification in the three studies noted, which were the only experiments that used body shape descriptors for identification purposes. The results showed an initial consideration of what the performance of the body shape description would be in human identification. However, both studies used mmW images, which are high quality images. Therefore, there is a gap in the literature where experimental studies that explore the performance of using the body shape description as a digital representation on images that simulate the CCTV images' quality is lacking. Consequently, the performance of using shape descriptors for person re-id has not been established.

2.3.3.4 Other body traits

Apart from the above-mentioned soft biometrics, there are other appearance-based traits that can be extracted at a distance (e.g. human height and weight). Estimating the height and weight is explored in soft biometrics field by either a numerical scale estimation or a list of categorical labels (i.e. very tall, tall and short).

For example, [58] improved an approach based on [59] for body-based measurement which predicted a score for the human silhouette height and weight. The improvement included predicting a scale height and weight from a single view and low-rate frames. After extracting the silhouette, they calculated the height of a subject from a 2-D single view frame thus:

$$O_H = \left(\frac{H_C}{V}\right) * (X_T - X_B) \tag{2.3}$$

where H_C is the predefined parameter of the camera height, and X_T and X_B are the top and bottom of the silhouette boundary, respectively. Considering that X_T and X_B are points in different planes, their projection is the vanishing point V. The distribution of the estimated height around the actual height is shown in Figure 2.9.

On the other hand, the researchers in [58] and [60] highlight that weight estimation methods are very limited, compared with height. In addition, the precise estimate of human weight is not always possible, mainly because weight is significantly affected by noise. To estimate weight, the subject silhouette was divided into head, torso and leg zones. Then, a further 12 features were calculated in addition to O_H, which represents the subject height. Figure 2.10 shows the 13 different regions that were computed for calculating weight. The proposed height and weight methods were implemented on a private dataset, consisting of 80 subjects. The height method yielded a 1.57 cm error and a 3.6 cm standard deviation, which outperforms the techniques presented in [61]–[64] in either the error rate or the frame rate. In terms of weight, the error was 4.66 kg.

In [65], the same methods proposed in [58] were integrated with facial shape feature and skin colour descriptor for the purpose of human identification. The schematic proposed system is shown in Figure 2.11. For each feature, a single descriptor was generated, combining all the information from several frames of the same subject, when they were in the camera's sight. Consequently, the matching process was conducted separately for each feature to generate the similarity scores. Then, the scores were fused using a number of fusion techniques, which included the sum rule, weighted sum, fuzzy logic, Bayesian and SVM.

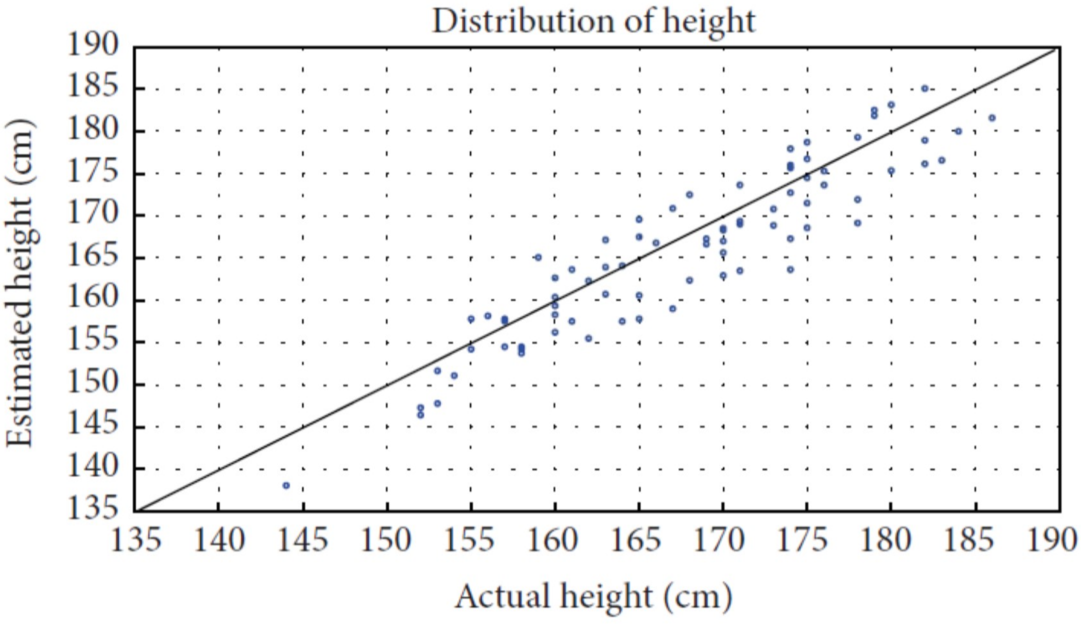


Figure 2.9: The distribution of the estimated height around the actual height using the method presented in [58]

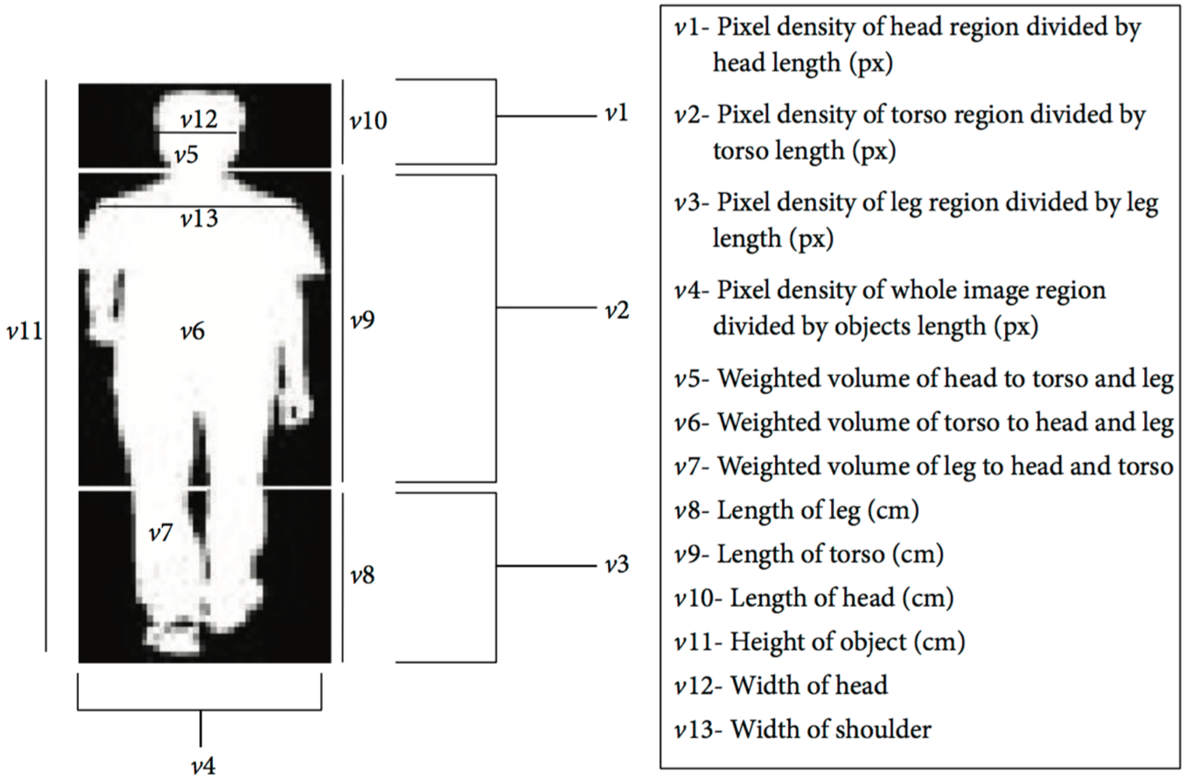

v12
v5
v13
v6
v7
v11
v10
v9
v8
v4
v1
v2
v3
v1- Pixel density of head region divided by head length (px)
v2- Pixel density of torso region divided by torso length (px)
v3- Pixel density of leg region divided by leg length (px)
v4- Pixel density of whole image region divided by objects length (px)
v5- Weighted volume of head to torso and leg
v6- Weighted volume of torso to head and leg
v7- Weighted volume of leg to head and torso
v8- Length of leg (cm)
v9- Length of torso (cm)
v10- Length of head (cm)
v11- Height of object (cm)
v12- Width of head
v13- Width of shoulder

Figure 2.10: Collections of features for estimating weight [58]

This experiment was conducted on Chokepoint dataset [66], which is considered the only publicly available face surveillance dataset (see [65]). The performance was evaluated in three scenarios, which included evaluating (a) individual features performance, (b) different feature combinations and (c) performance of multiple score fusion. The results of the single features showed that the height performance outperformed the weight and skin colour descriptors' performance by a 30% first rank accuracy rate for height and 14.2% for weight and skin colour. However, the result of the facial shape descriptor is double that of height with a 64.3% accuracy rate at first rank. In terms of combining different features, the results showed that merging facial shape, height and weight descriptors outperforms other combinations. Figure 2.12 shows the performance of different descriptor combinations as well as the performance of different score fusion techniques.

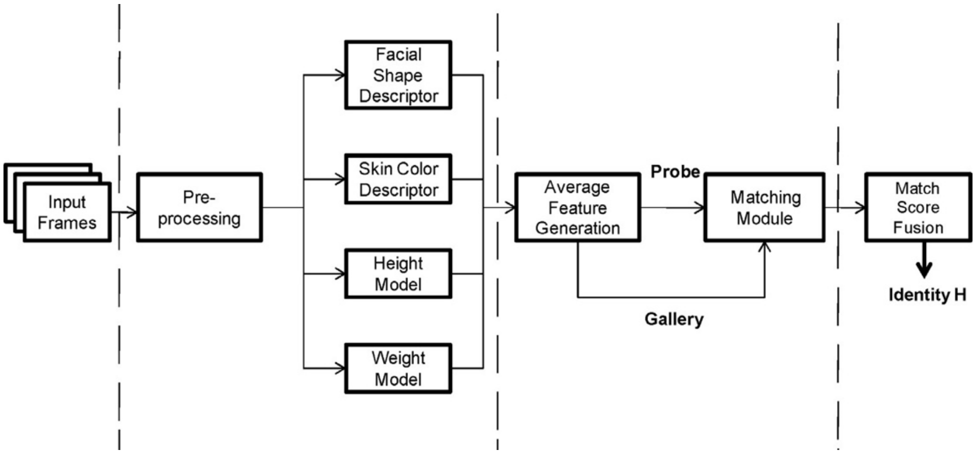


Figure 2.11: Facial shape, skin colour, height and weight features fusion system framework proposed in [65]

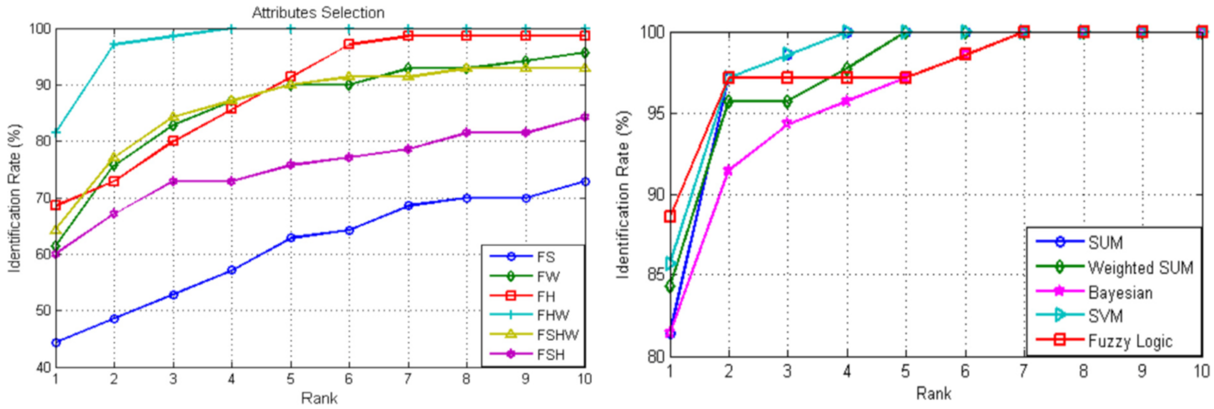


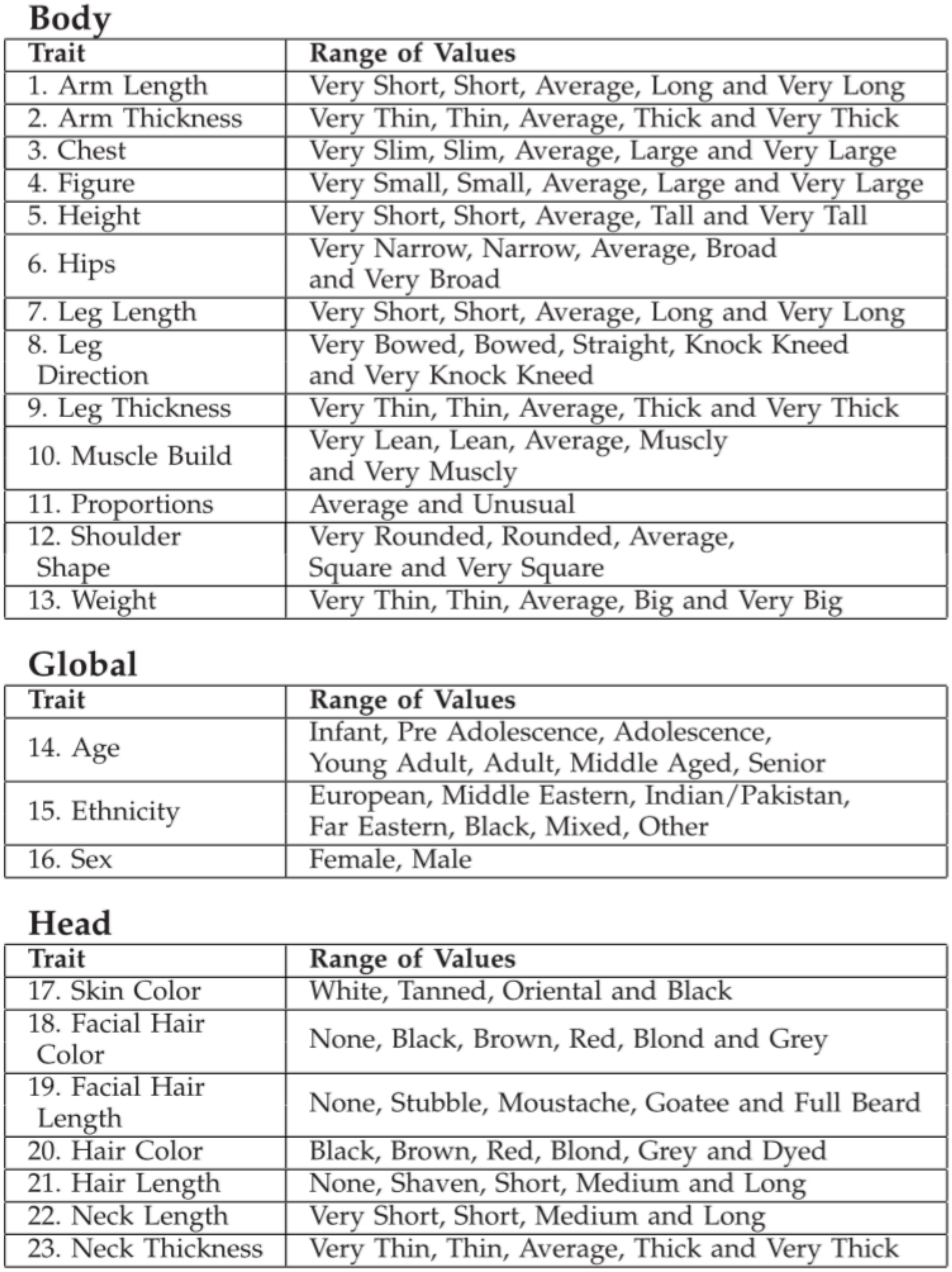
Figure 2.12: Left, combining different features, F = Facial, S = Skin colour, W = Weight, H = Height. Right, the performance of different score fusion techniques [65]

Another way of estimating a numerical height value is through camera calibration and its intrinsic parameters. Several studies have introduced and improved such approaches (see [67], [68]). On the other hand, the human height, weight and other soft biometrics can be measured using a categorical label. For example, [69] examined a number of soft biometrics in short, medium and far distances between the subject and the camera. Each trait was manually annotated using one of the labels in Table 2.5. The annotation process involved asking an annotator; the annotations were subsequently fused with the face recognition system, which showed comparable performance enhancement.

In [70], a number of soft biometrics were automatically extracted from a single shot (i.e. one frame), which included height, shoulders and hips' width, arm's length, body complexion and hair colour. For each soft trait, there are several categorical labels (see Table 2.6). The dataset was used in [70],was Southampton Multi-Biometric Tunnel Dataset [71], containing 222 subjects. The height was extracted by measuring the distance between the top and bottom points of the *y* axis of the extracted subject silhouette. To evaluate performance, the ground truths were manually extracted by using one height label, which was either short, average or tall; these were then compared with automatic category of the same subject. SVM was used as a classifier to determine whether each subject was short, average or tall or otherwise. The performance evaluation showed promising results; the accuracy rate for height was 94.6%, with 24 (out of 222) wrongly classified results. The categorisation of height outperforms that of all the other soft traits included in this study.

Generally, the studies presented showed that height and weight features are extracted based on the body shape of the subject. The literature showed that these features are represented as a scale number or by assigning a categorical label (i.e., average or medium). Most of the studies focus on improving the feature extraction methodology more than exploring the influence of using these features as identifiers for person re-identification. No evidence was found to indicate the prior use of height and weight as subject representations in any person re-id applications, individually or along with other features. A number of surveys, however, have been conducted on soft biometrics (e.g., [60], [72], [73]).

Table 2.5: Soft biometrics and their categorical labels (see [69])



Body

Trait	Range of Values
1. Arm Length	Very Short, Short, Average, Long and Very Long
2. Arm Thickness	Very Thin, Thin, Average, Thick and Very Thick
3. Chest	Very Slim, Slim, Average, Large and Very Large
4. Figure	Very Small, Small, Average, Large and Very Large
5. Height	Very Short, Short, Average, Tall and Very Tall
6. Hips	Very Narrow, Narrow, Average, Broad and Very Broad
7. Leg Length	Very Short, Short, Average, Long and Very Long
8. Leg Direction	Very Bowed, Bowed, Straight, Knock Kneed and Very Knock Kneed
9. Leg Thickness	Very Thin, Thin, Average, Thick and Very Thick
10. Muscle Build	Very Lean, Lean, Average, Muscly and Very Muscly
11. Proportions	Average and Unusual
12. Shoulder Shape	Very Rounded, Rounded, Average, Square and Very Square
13. Weight	Very Thin, Thin, Average, Big and Very Big

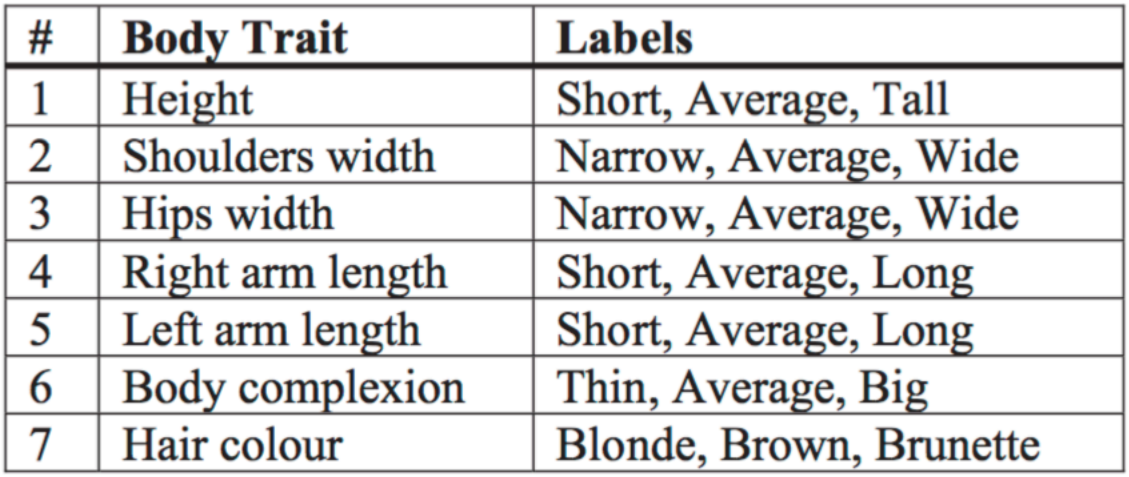
Global

Trait	Range of Values
14. Age	Infant, Pre Adolescence, Adolescence, Young Adult, Adult, Middle Aged, Senior
15. Ethnicity	European, Middle Eastern, Indian/Pakistan, Far Eastern, Black, Mixed, Other
16. Sex	Female, Male

Head

Trait	Range of Values
17. Skin Color	White, Tanned, Oriental and Black
18. Facial Hair Color	None, Black, Brown, Red, Blond and Grey
19. Facial Hair Length	None, Stubble, Moustache, Goatee and Full Beard
20. Hair Color	Black, Brown, Red, Blond, Grey and Dyed
21. Hair Length	None, Shaven, Short, Medium and Long
22. Neck Length	Very Short, Short, Medium and Long
23. Neck Thickness	Very Thin, Thin, Average, Thick and Very Thick

Table 2.6: Soft traits with their categorical labels [70]



#	Body Trait	Labels
1	Height	Short, Average, Tall
2	Shoulders width	Narrow, Average, Wide
3	Hips width	Narrow, Average, Wide
4	Right arm length	Short, Average, Long
5	Left arm length	Short, Average, Long
6	Body complexion	Thin, Average, Big
7	Hair colour	Blonde, Brown, Brunette

2.3.3.5 Behavioural cues

Human gait is one of the most common behavioural biometrics. It is a valuable feature of person re-id, mainly because a deliberate change to it will look unnatural [25]. Although the walking speed, outfit type and mood are likely to affect gait, their influence on gait is consistent over short-term re-id processes [25]. Gait techniques, are challenging traits because analysing gait requires a clear view side of a subject for at least one or two steps [25].

In [74], the gait features were integrated with appearance-based features. The gait feature was extracted using Gait Energy Image (GEI), by drawing on the silhouettes' sequences from different angles separately. The dataset used in this study was CASIA Dataset B, which was originally a gait dataset. However, this dataset has been used recently for person re-id applications. The image sequences of this dataset were collected by 11 overlapped cameras with different viewing angles. Figure 2.13 shows the GEI of the sequences from the 11 views. GEI was combined with HSV, as a colour descriptor, and Gabor filter, as a texture descriptor. Two different frameworks were used in this experiment. The first framework included score-level fusion, which fused the distances originating from the gait and appearance features. The second framework included feature-level fusion, which fused the features before obtaining the distances. Figure 2.14 shows the different flows of the frameworks. The performance of the method presented in [74] was implemented on two frameworks over the three scenarios and compared with a number of other similar studies, such as SDALF [75], ELF, ITML, and LMNN [76].

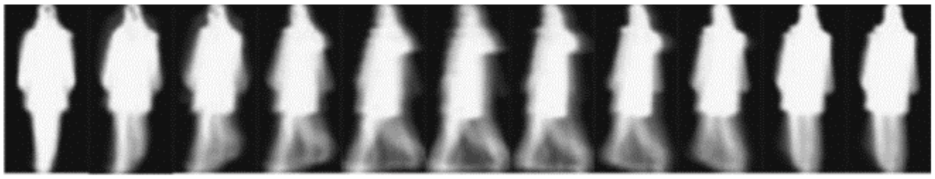


Figure 2.13: Gait Energy Image (GEI) of sequences of silhouettes from 11 different angles [74]

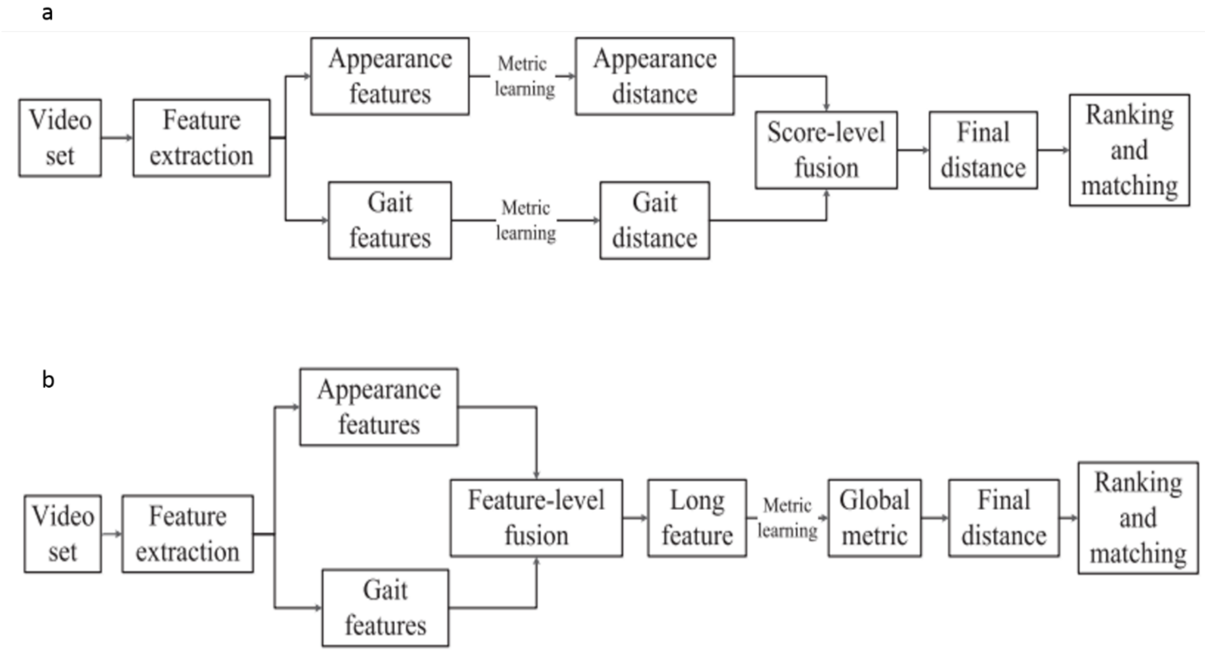


Figure 2.14: (a) Score-level fusion framework; (b) Feature-level fusion framework [74]

Several experiments were also conducted, using closed set and open set person re-id approaches. The closed set means that the gallery set size is fixed, whereas the open set person re-id means that the gallery set size evolves by time simulating the real scenario of person re-id. Figure 2.15 and Figure 2.16 show the CMC curves of the experiments conducted on closed and open set respectively and with the three scenarios namely bag, clothes and normal.

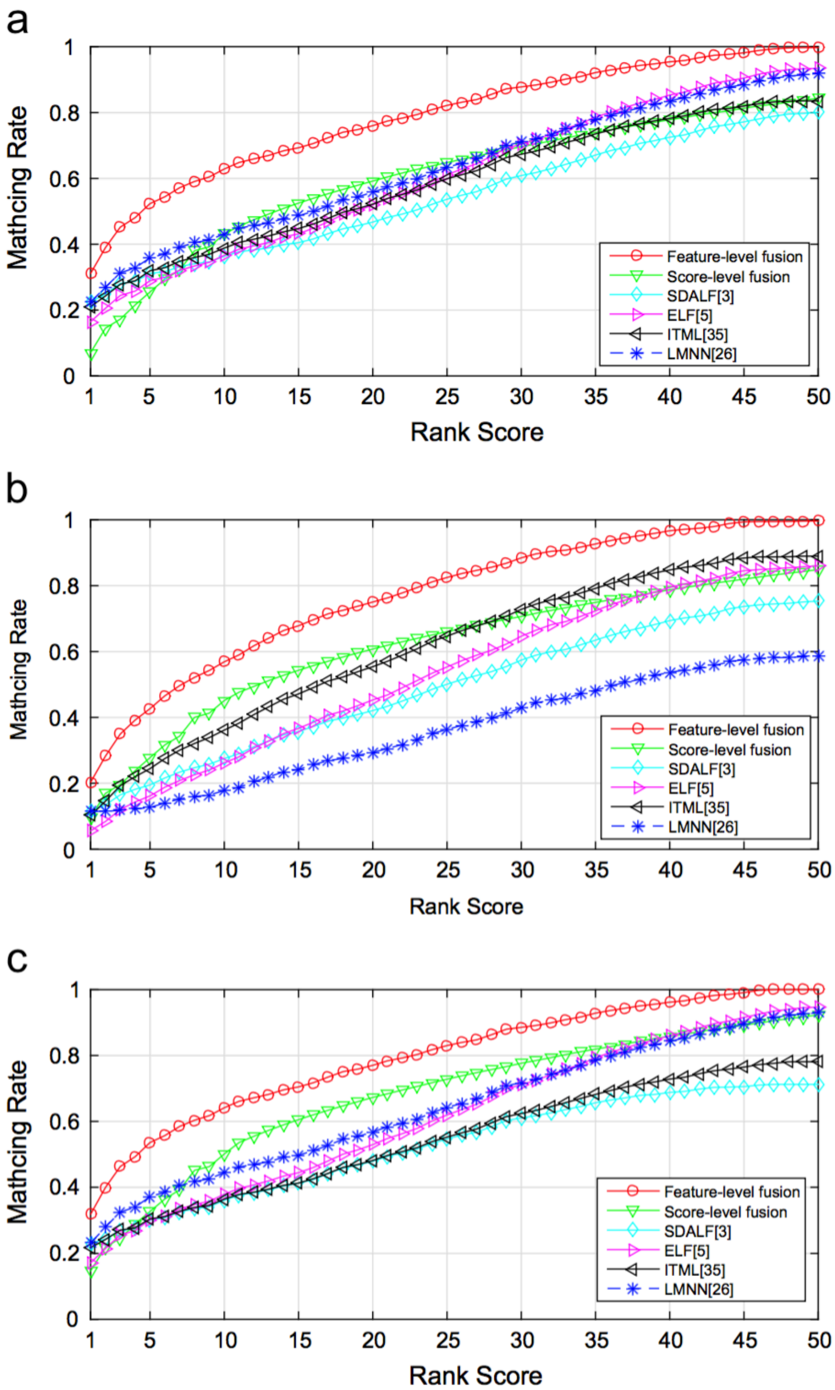


Figure 2.15: (a), (b) and (c) are closed set person re-id CMC curves of the implementation of feature level fusion and score level fusion, compared with a number of other techniques on 'bag-bag', 'clothes-clothes' and 'normal-normal' scenarios, respectively [74].

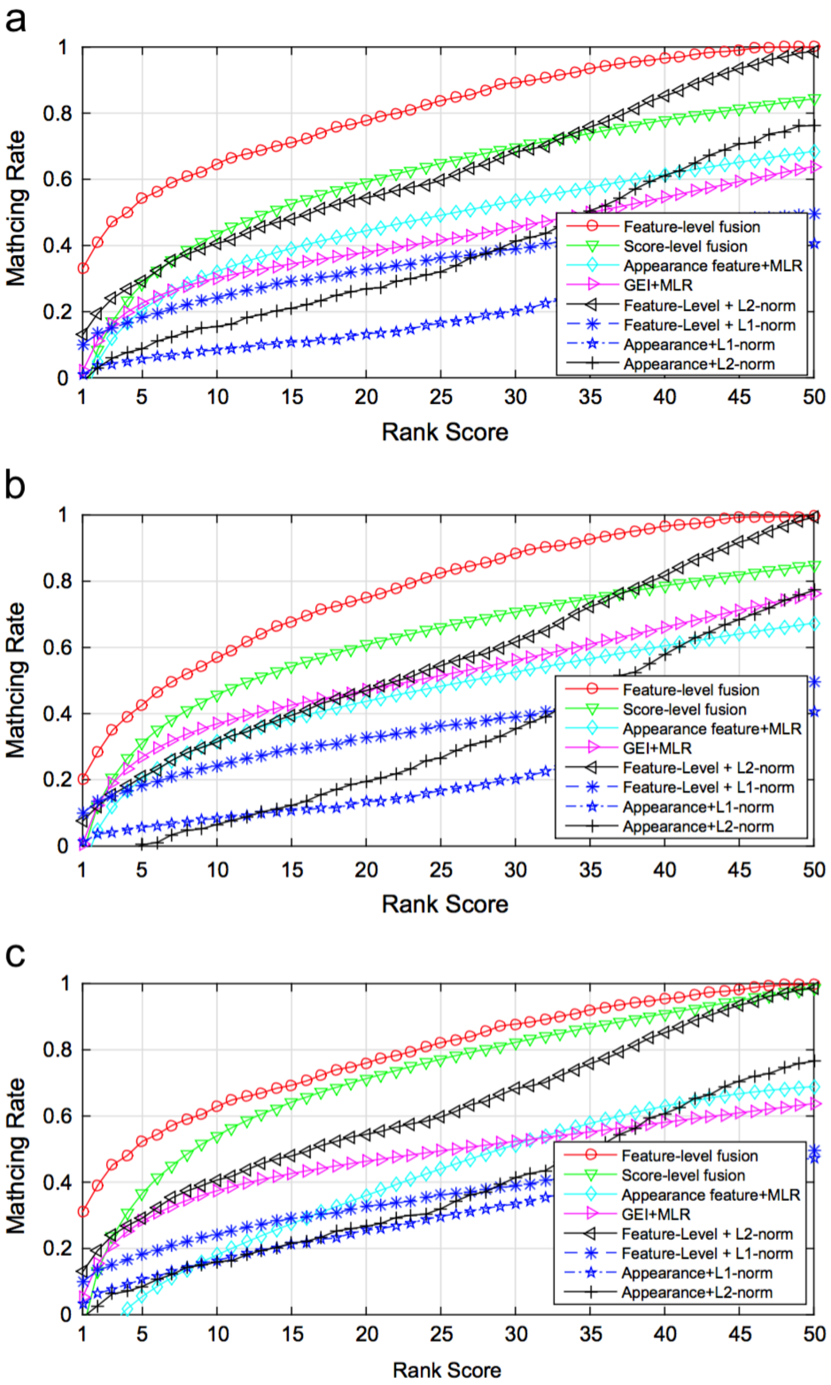


Figure 2.16: (a), (b) and (c) are the open set person re-id CMC curves of the implementation of feature level fusion and score level fusion, compared with a number of other techniques on 'bag-bag', 'clothes-clothes' and 'normal-normal' scenarios, respectively [74].

In [77], the gait feature was also extracted using GEI. However, the distances were matched, using improved version of multi-dimensional DTW, called Multi-Dimensional Time Shift Dynamic Time Wrapping (MDTS-DTW). Instead of matching the whole two sequences in DTW, MDTS-DTW was developed to iteratively and partially match the sequences. Figure 2.17 shows an illustration of MDTS-DTW, where X^p is the probe sequence and X^g is the gallery sequence. In this method, the distance between $X^p$and X^g is computed at every time point Δt. This experiment was conducted on the PRID2011 and iLIDS-VID datasets. iLIDS-VID dataset consists of 300 subjects recorded from two non-overlapped cameras in an open public area. The performance of this method was compared with other person re-id methods, known as SS-SDALF, MS-SDALF, ISR, eSDC and RDL. The performances shown in Table 2.7.

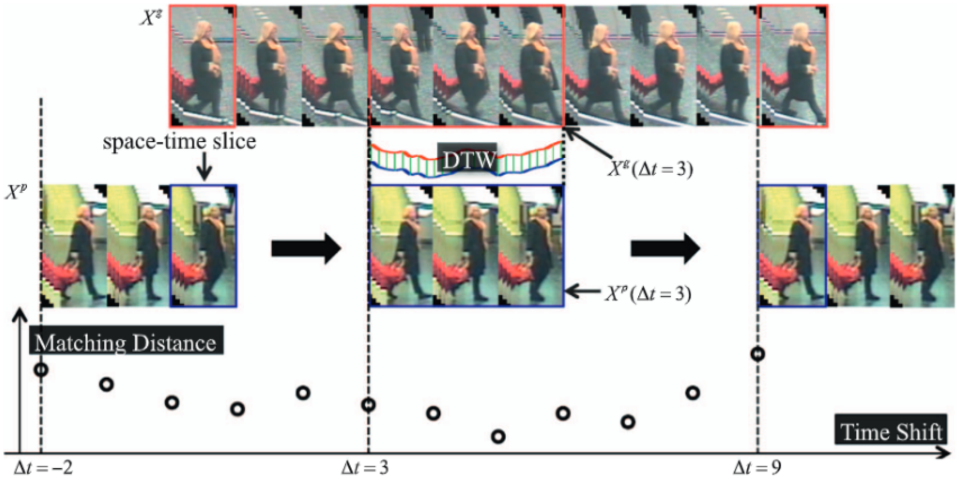
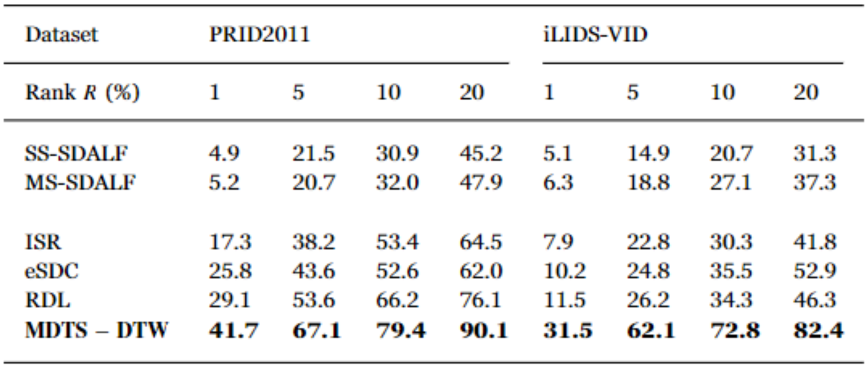


Figure 2.17: Overview of MDTS-DTW, proposed in [77]

Table 2.7: Comparing the performance of MDTS-DTW with other re-id methods [77] (in percentage)



Dataset	PRID2011				iLIDS-VID			
Rank R (%)	1	5	10	20	1	5	10	20
SS-SDALF	4.9	21.5	30.9	45.2	5.1	14.9	20.7	31.3
MS-SDALF	5.2	20.7	32.0	47.9	6.3	18.8	27.1	37.3
ISR	17.3	38.2	53.4	64.5	7.9	22.8	30.3	41.8
eSDC	25.8	43.6	52.6	62.0	10.2	24.8	35.5	52.9
RDL	29.1	53.6	66.2	76.1	11.5	26.2	34.3	46.3
MDTS – DTW	**41.7**	**67.1**	**79.4**	**90.1**	**31.5**	**62.1**	**72.8**	**82.4**

Behavioural cues, gait features especially, are among the most used and investigated features in person re-id applications. The convenience of extracting these features is one reason for their popularity, as they can be extracted at a distance. Moreover, their performance efficiency is more accurate compared to colour, texture, and clothes features. Gait feature is one of the features that can be extracted based on the body shape, along with height, weight, and body shape description. Gait is the only trait out of all of these that is exploited for person re-id application.

2.3.3.6 Temporal cues

Entry, exit time, locations (i.e. positions), velocities and transition times between cameras are examples of temporal cues, used as enhancement for human re-id by either reducing the number of correspondences (limiting the size of the gallery set) or extracting the features for re-id.

Assuming that humans in public areas follow the same path, the researchers in [78] and [79] developed a tracking system that exploits the entry and exit locations, using the velocities and the transition times to establish a link between the subjects' inter-cameras. The system used camera topology rather than camera calibration. In addition, [80] used people trajectories to specify the regions of interest in the non-observed areas within the FOV of the cameras. This helped limit the number of the paths that the subjects might take, constraining the number of areas, in which the subjects would reappear. Figure 2.18 shows subjects' trajectories evaluation results by detecting the existence and the absence of people on two camera frame streams.

In general, less focus was placed on exploring the performance of temporal cues (i.e., entry and exit time, locations, velocities, transition times, and trajectories) in identification applications. Exploiting temporal cues for person re-id applications and their influence on performance can be a future topic of study to enhance the field.

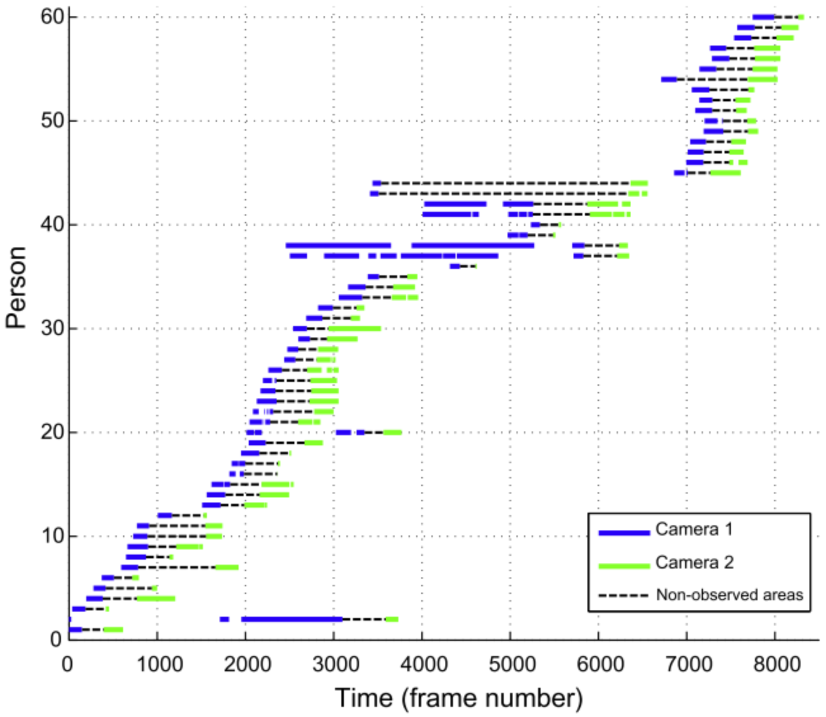


Figure 2.18: Evaluating people's trajectories from Camera 1 (Blue) and Camera 2 (Green), where the dotted lines are the non-observed areas [80]

2.3.4 Distance Metrics and Classification for Similarity Estimation

In the previews section 2.3.3, the focus was on the second stage of person re-id, shown in Figure 2.1. It discussed what and how to represent and extract unique features from people who are moving in the FOV of the non-overlapping camera networks. This Section focuses on the third stage of person re-id, which includes the methods and techniques that have been used to compare the representations and match the most similar correspondences.

Number of distance metrics have been used for different types of biometric applications, such as the sum of the quadratic distances [81], the sum of the absolute differences[82], [83], the correlation coefficients [84], the Bhattacharyya coefficients [85] and the Euclidean distances. However, these metrics share non-flexibility, which means that they deal with all the fed features equally, discarding the less useful features. In re-id, however, non-flexibility may lead to considerable limitations [25].

Therefore, Mahalanobis distance metric is a commonly used distance metrics in person re-id applications [25]. The advantage of Mahalanobis metrics is that it calculates and considers the correlations between the provided features' vectors. Another popular metric learning method in person re-id is KISSME [86] which was based on Mahalanobis distance metric. In KISSME, a likelihood ratio test was formulated to determine whether a pair of vectors are similar or not. Large Margin Nearest Neighbour (LMNN) [76] is another learning method, which sets up a boundary for the match pair's neighbour, assuming that they mostly belong to the same class. Inter-view Quadratic Discriminant Analysis (XQAD) [87] is another distance metric which is wieldy used for person re-id. XQAD learn a low dimensional subspaces using inter-view quadratic discriminant analysis. In [88], the optimal matching measures between a prop and gallery images is learned a Relative Distance Comparison model (RCD). RCD maximises the probability of the true match pairs with a relatively smaller distance in a soft discriminant manner. Others also use learning methods such as Support Vector Machine (SVM), k-nearest neighbour KNN [89], structural SVM [90] and AdaBoost [91].

2.3.5 Deep Learning Person Re-id Systems

The recent direction of person re-id inclines towards deep learning methods instead of the general framework, shown in

Figure 2.1. Conventional Neural Networks (CNN) as deep learning models, were first used in re-id in 2014 in [83] and [93] (see [94]). These models teach deep learning features and classify subjects according to an end-to-end model. One major issue with CNN models in re-id is the lack of training data; this is mainly because datasets mostly provide only a few numbers of images per identity. Therefore, re-id deep learning models focus on the Siamese model [95], which only uses double or triple images as input.

In [93], Siamese CNN (SCNN) image input was horizontally divided into three parts, including head, body and leg. The parts underwent two convolution layers, two max pooling layers and a full connected layer. The fully connected layer fused with the output to generate a vector for the original input image. The similarity between the vectors was computed by cosine distance. Figure 2.19 shows the SCNN used in this experiment. The experiment of using the Siamese CNN model for person re-id was conducted in two manners: (a) training and testing, using the same dataset (e.g. VIPeR), which is actually a Single dataset and (b) using Cross datasets, using CUHK Campus for training and VIPeR for testing. The CMC of the performance of SCNN on Single and Cross datasets can be seen in

Figure 2.20. The performance of SCNN is compared with a number of state-of-arts of the representation and matching methods as shown in Table 2.8.

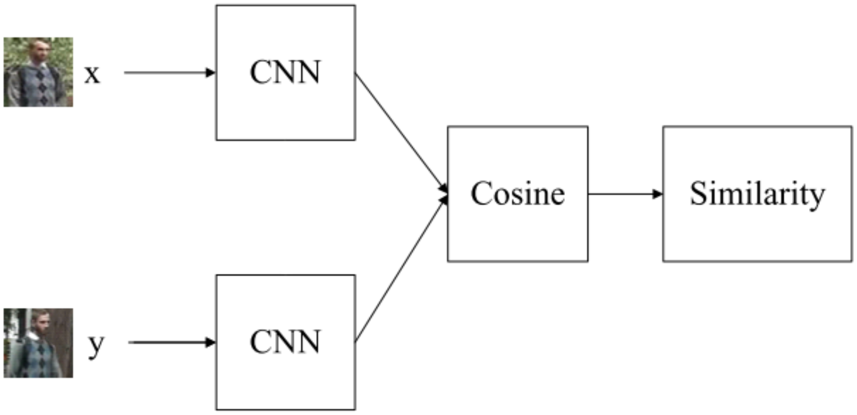


Figure 2.19: Siamese SCNN model presented in [93]

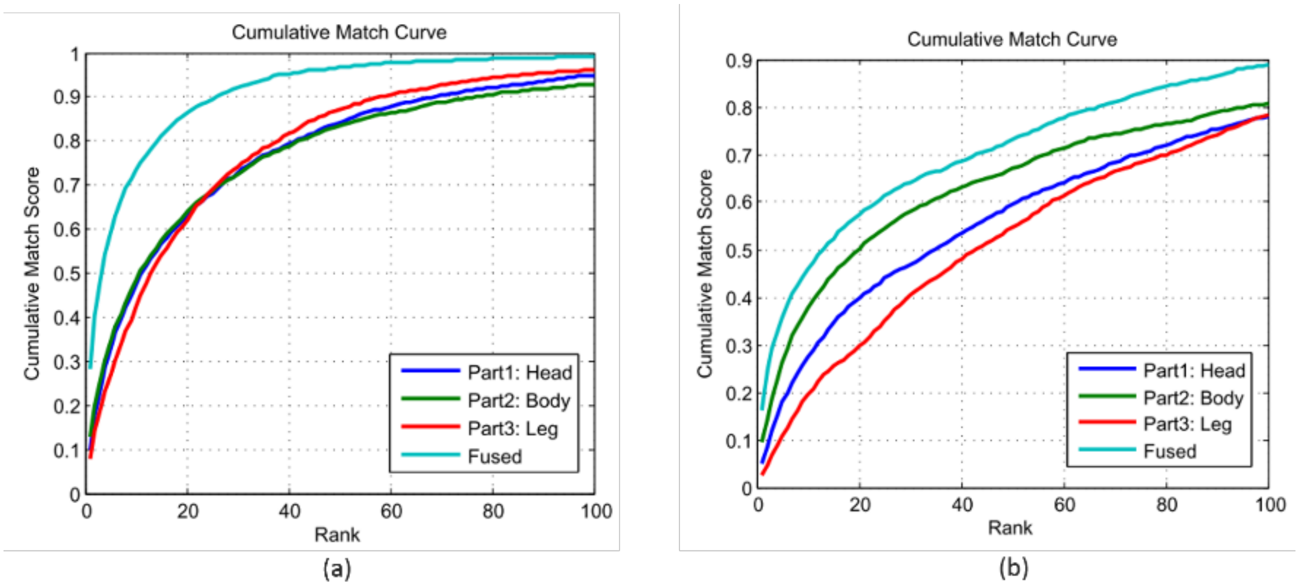
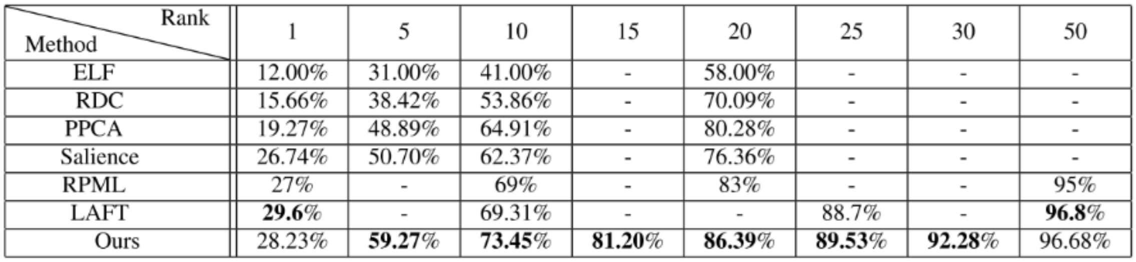


Figure 2.20: (a) Siamese SCNN performance on single dataset; (b) Siamese SCNN performance on multiple dataset [93]

Table 2.8: Performance comparison of several re-id methods and proposed methods [93]



Rank Method	1	5	10	15	20	25	30	50
ELF	12.00%	31.00%	41.00%	-	58.00%	-	-	-
RDC	15.66%	38.42%	53.86%	-	70.09%	-	-	-
PPCA	19.27%	48.89%	64.91%	-	80.28%	-	-	-
Salience	26.74%	50.70%	62.37%	-	76.36%	-	-	-
RPML	27%	-	69%	-	83%	-	-	95%
LAFT	**29.6%**	-	69.31%	-	-	88.7%	-	**96.8%**
Ours	28.23%	**59.27%**	**73.45%**	**81.20%**	**86.39%**	**89.53%**	**92.28%**	96.68%

Another experiment that used SCNN for person re-id is presented in [92]. The difference between this experiment and the previous experiment is that the responses of the two images from the convolution layers are multiplied, using a patch-matching layer.

In [96], the Siamese model improved cross-input neighbourhood variations by comparing the features of the first input image with all the close locations in the other input image. Then the output of the convolution layers were subtracted, using patch-matching layers to find the similarity.

In [97], long short-term memory (LSTM) models are integrated into the Siamese network. In LSTM, the image segments are processed sequentially in order to memorise the spatial connections, which maximise the discrimination of the deep features.

2.4 Body Segmentation

Methods and techniques that have been developed to arithmetically describe people in public places tend to do so using the entire body or part of it. Dividing the human body into respective parts can be achieved through number of approaches. The first way involves partitioning the bounding box of the subject into fixed proportions [98]. In [99], the body was divided into three regions (i.e. head, torso and legs), using two asymmetrical axes. The first axis that divided the head and torso was determined by finding the maximum differences between the numbers of the pixels in two rectangles. The second axis that divided the torso and the leg parts was identified by the maximum colour differences. The torso section and leg sections were vertically partitioned into left and right parts by weighting the pixels in each side. Figure 2.21 shows the proposed approach of body partitioning.

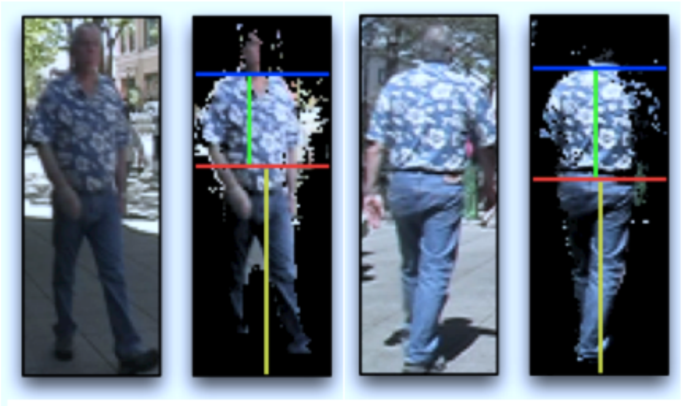


Figure 2.21: Body partitioning using the proposed method in [99]

The other way of segmenting the body is a spatial-temporal segmentation method that was developed in [26] to partition the subject's silhouette using their edges. This method is invariant to clothes wrinkles and lighting variations. The human detection HOG technique was also used for the detection of body parts in [100] and [101] as shown in Figure 2.22. Similarly in human detection, the HOG method is trained on positive and negative samples of body parts.

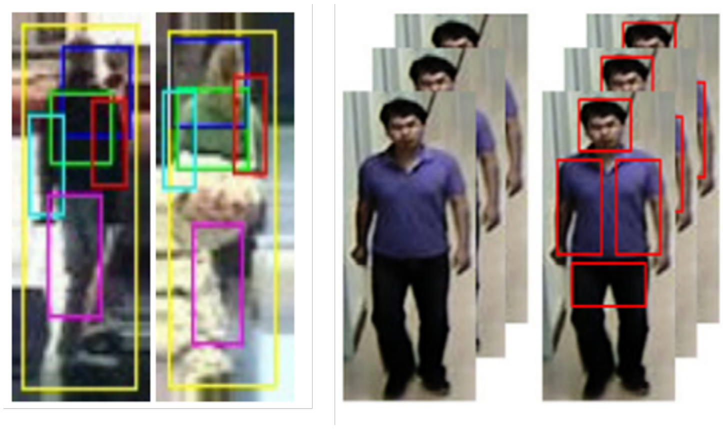


Figure 2.22: Examples of using HOG method for segmenting the body in [92] and [93]

2.5 Summary and Discussion

This research involves investigating the accuracy of using body shape for the purpose of person re-id; it also intends to probe the stability level of body segment shapes. Because viewing angle plays a major role in influencing the individual representation of a subject, the impact of this factor upon the shape of the body parts is also examined. This experimental study involves a number of research areas, such as biometric systems, feature extraction and representation, similarity estimation methods, and body partitioning techniques. This literature review chapter addresses each of them.

The first section of this chapter outlines the biometric systems categories; it focuses on person re-id biometric systems. In the second section, a detailed survey on person re-id is presented, which includes evaluation metrics, representation techniques, distance metric methods, and deep learning methods in person re-id. The final section is concentrated on the number of body segmentation applications used in the literature.

This literature review illustrates that the person re-id field is an underdeveloped research area, compared with the face and fingerprints verification methods and the enormous investments that they receive (see section 2.3.1 for performance challenges). As the first publication that independently worked on person re-id was [26] in 2006 and deep learning was only applied to it on 2014 (see [94])

A single situation for developing a person re-id system that addresses all the challenges does not yet exist. Most of the systems noted in the literature that use representation and similarity estimation tend to address one aspect of the problem, leading to a lack of end-to-end systems. The recent trend of person re-id towards exploiting deep learning methods can be interpreted as an attempt to address end-to-end systems.

In general, the main characteristic of person re-id is the evolution of the gallery set, where the number of the enrolled subjects is not fixed. Therefore, the enrolment process is a fundamental step in a person re-id application. However, little attention has been paid to this stage of implementation in the literature. In addition, there are many datasets that were especially designed for person re-id implementation. However, every dataset was created for a certain investigation, meaning there is a lack of datasets that are publicly available and can provide a wide range of appearance-based features' ground truths.

In terms of the person re-id approach, which involves feature representation and similarity estimation, studies on the state-of-art systems primarily focused on features such as colour, texture, and gait, while integrating clothing attributes with other person re-id traits received less attention. In addition, the body shape-related features, such as body shape description, height, and weight, have not been explored in any person re-id systems. These features require experimental study to examine the role of human body shape in person re-id and to inspect the effectiveness of body segments in person re-id with respect to viewing angle. Therefore, focus has been placed on these aspects in this research study.

Chapter 3

Experimental Framework

3.1 Introduction

In this research, the human body shape is defined using the silhouette of the body. A silhouette frame in image processing is the image frame of a person that is represented as a solid shape of white colour, where the shape contour matches the outline of the subject's shape. The inside area of the silhouette is featureless, and the background surrounding the subject is black. The silhouette frame can be obtained from the original, full-coloured image frame by detecting the subject in the image frame and assigning all the pixels belonging to the subject's body shape a white colour. All other pixels in the image are converted to black as the background colour. A silhouette frame example can be found in Figure 3.1.

Publicly available datasets were browsed to locate a dataset with images that met four main criteria: (a) the same subjects recorded from multiple camera views, i.e., the subjects' image sequences were recorded from different viewing angles; (b) tracking sequences recorded of the same subject, i.e., image sequences that recorded a subject from one point in time to another, providing followed frames of the subject; (c) original, full-coloured image sequences with their corresponding silhouette image sequence frames; and d) the same subjects recorded with several different appearances. This chapter provides a review of existing person re-id datasets and describes selection of the target dataset based on the above-mentioned criteria.

This chapter also presents the calculation of the anatomical average value of the length parameters of non-overlapping parts of the human body, comprising *Head and Neck*, *Shoulders*, *Middle,* and *Lower*. The length parameters will be subsequently used to arithmetically segment the human body silhouettes into these four non-overlapped sections.

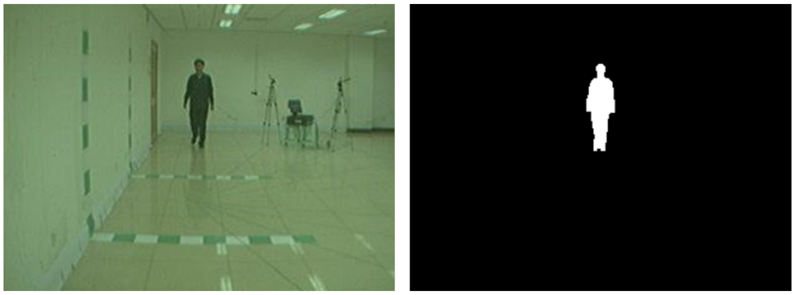


Figure 3.1: Original frame on left, corresponding silhouette frame on right

This chapter is organised as follows: Section 3.2 presents the proposed framework for the shape-based person re-id system and a discussion on its stages; section 3.3 introduces publicly available person re-id datasets and the rationale behind selecting CASIA Dataset B as the examined dataset for this research. Section 3.4 provides a description of the proposed silhouette segmentation methodology, including the anatomical average length of each body segment based on multiple anthropological studies; in addition, the arithmetic operation that algorithmically divides the human body silhouette into the four suggested segments is illustrated.

3.2 Proposed System

The proposed shape-based person re-id system investigates the effectiveness of shape descriptors of the body and body parts (i.e., segments) of humans for person re-id. Hence, the first stage involves pre-processing the silhouette sequences by dividing the silhouettes into the corresponding segments, which is called the *segmentation process*. This leads to exploring the power of each segment separately, followed by discovering the factors that negatively affect the discrimination of each body segment. The segmentation process is applied mainly by finding the smallest bounding box around the subject's silhouette frame. Then, the anatomical average length of each body segment is applied to a arithmetic operation that arithmetically and automatically segments all the silhouettes in the selected dataset into the proposed body sections. Body segmentations are explained in Section 3.4.

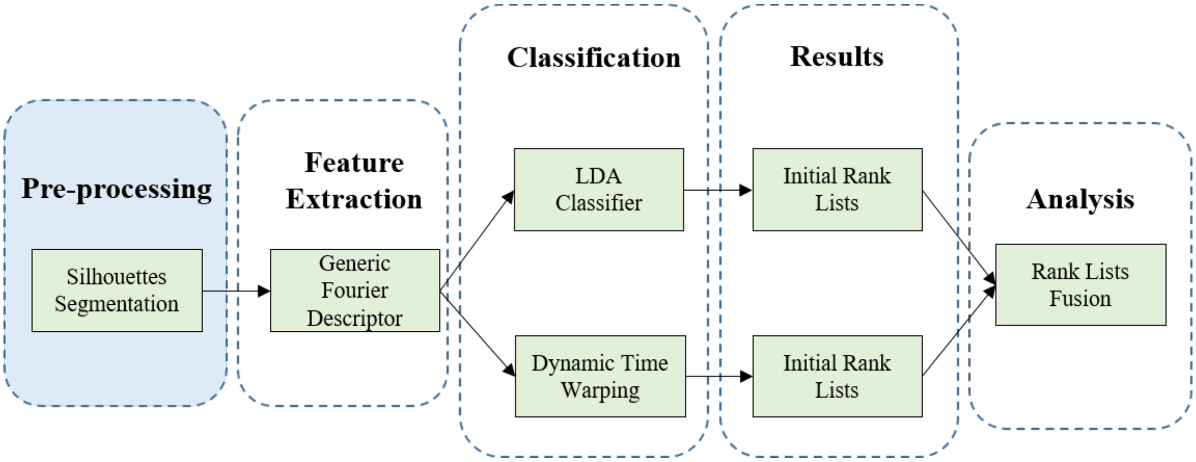


Figure 3.2: Proposed shape-based person re-id system framework

The second stage of the proposed system involves the feature extraction process; the feature explored in this research is a shape descriptor of each body segment. The Generic Fourier Descriptor (GFD) shape descriptor is extracted from each segment individually. Shape descriptors in general are divided into contour-based shape descriptors and region-based shape descriptors. Contour-based descriptors only extract information located at the boundary; the interior content of the shape cannot be used. Therefore, contour-based descriptor applications are limited. The region-based shape descriptors are extracted utilising all pixel information in the contour and inside a shape region. Region-based shape descriptors are implemented in more applications than contour-based descriptors. The performance of the GFD was compared in [102] with the common contour-based and region-based shape descriptors. The results showed that GFD outperformed the other two techniques. The details implementation of the GFD can be found in 4.2. In this stage, the data are split into training and test sets, as each subject has two different silhouette sequences, one for each set.

The next stage of the proposed system is the assessment of the feasibility of the shape descriptors of the proposed body segments (i.e., classification stage). The system is designed for the assessment to be conducted through two approaches, image-based and video-based, as any person re-id application is designed as either an image-based system or a video-based system [103]. In the image-based system, the shape descriptors of the segments are classified using Linear Discernment Analysis (LDA) [104], resulting in an initial rank list for each body segment frame. LDA mainly maximises the distances between the mean of the classes and minimises the variation within one class's samples. LDA explanation can be found in 4.2.3. This part of the system examines the static variations of the segments, as the classification process is done on a frame-by-frame basis.

For the video-based system, multidimensional Dynamic Time Wrapping (DTW) [105] is applied to the GFD shape descriptor of each segment. DTW calculates the similarity between two temporal sequences. It returns a distance scalar of the two sequences' optimal alignment. Detailed multidimensional-DTW can be found in 5.2. The shape descriptors of all the frames in one sequence in the training set are wrapped with all the sequences in the test set to find the most similar sequence. Through this process, the dynamic features of the shape descriptors within a sequence are examined, and then one rank list for each sequence is generated.

The final stage of this system involves applying a proposed rank lists fusion method. The fusion method combines the image-based generated rank lists so that the lists generated by all frames in each sequence are replaced by one rank list. To develop the proposed rank list, the indices of each identity in the initial rank lists are counted, and the indices of each identity from each initial rank list are added together, and then the identity with the least total indices is placed in the frontal location of the new fused rank list.

The methodologies and the experimental results of each stage are explained in detail in the next chapters. This chapter focuses on the pre-processing step, as highlighted in Figure 3.2.

3.3 Dataset Description

Finding a target dataset containing the characteristics necessary to answer the research questions is a fundamental matter in this study. In order to answer the research questions that interrogate the performance of the body shape descriptor, the power of each segment of the body, the factors influencing performance, and the silhouette shape of the human body are exploited. To facilitate this, a dataset that can provide the silhouettes' frames as well as multiple viewpoint sequences that record the same subject from different viewing angles is needed.

Another goal of this study is to explore the effectiveness of the body segments shape descriptors while the subjects are in different scenarios, such as appearing in different types of clothing in the same scene. Thus, a dataset that recorded the same subject in different scenarios is required.

Exploring the dynamic variation of the shape within a sequence is another investigation involved in this research. Hence, it is necessary to choose a dataset that provides tracking sequences. A tracking sequence is a recording of a subject from one point in time to another, including frames that follow the subject. There are, in fact, a number of datasets that provide these so-called *tracklets*, which are collections of images of the same subject from different points in time. For example, a tracklet can contain an image from each minute that the subject is in the camera's FOV. The tracking sequence is a record of the subject for the entire length of time it is present in a camera's FOV with a certain frame rate. The dynamic variation can be extracted from the tracking sequence, but missing frames may lead to performance failure.

In the literature, there are many publicly available datasets that have been widely used in person re-id research (e.g., VIPeR [106], iLIDS [107], CAVIAR4ReID [108], CASIA C [109], CASIA D [110], CUHK01 [111], V47 [112], and PKU-Reid [113]). However, these datasets, in particular, include a limited number of viewpoints (camera angles). On the other hand, GRID [114], 3DPeS [115], CUHK02 [116], RAiD [117], Market1501 [30], PRW [118], Large Scale Person search [119], DukeMTMC [120], Airport [121], and MSMT17 [122] provide multiple camera angles; however, they only offer a single or multiple shots rather than a sequence of followed frames.

Table 3.1 lists and summarises person re-id datasets that provide tracking sequences. A *Provides Silhouettes* column was added to the table to indicate whether the dataset offers the silhouettes' sequences corresponding to the original sequences because this study utilises the silhouettes to investigate the silhouette shape.

From

Table 3.1, it is clear that CASIA Dataset B and SAIVT-Softbio are the only datasets that provide three of the desirable characteristics for this research, namely, multiple camera views, tracking sequences, and silhouettes. However, after having assessed CASIA Dataset B and SAIVT-Softbio, it became clear that SAIVT-Softbio silhouettes contain extreme noise. Therefore, since CASIA Dataset B is the only dataset that provides records of the same subjects under three different scenarios, it is the selected dataset for this research, which is discussed in detail in the next subsection.

3.3.1 CASIA Dataset B

CASIA Dataset B [123] was captured by the National Laboratory of Pattern Recognition, China, in 2006. The dataset includes video recordings of 124 subjects captured in an indoor environment. The subject walking in a straight line was captured in three different scenarios: a) normal, b) wearing heavy clothes, and c) carrying a bag. Figure 3.3 shows the subjects' appearance in the three scenarios. In each scenario, each subject was recorded from 11 angles. Figure 3.4 shows the point of view of each angle. For each scenario, the subject was asked to repeat the walk to be recorded under the same conditions to provide two different sequences, as one of the sequences is used for the training and the other is used for the test. This dataset also provides silhouette images corresponding to the original video sequences. The silhouettes are stored in a PNG format of 320 x 240 pixels. In this research, and in each scenario, two different sequences are considered from each angle; the total number of subjects used for this study was 124, captured from 11 angles in three scenarios. Two different sequences per subject are included, bringing the number of sequences for investigation to 8,184, with an average of 68 frames per sequence.

Table 3.1: Reviewing person re-id datasets that provide tracking sequences

Name	Subjects	Views	Scene	Silhouettes
ETH1, 2, 3 [124]	85, 35, 28	1	Moving camera with a small viewpoint variance	No
PRID2011 [125]	934	2	Only 200 subjects appear in both cameras	No
WARD [126]	70	3	Non-overlapping cameras	No
SAIVT-Softbio [127]	152	8	Uncontrolled environment	Yes
iLIDS-VID [128]	300	2	Heavy occlusion	No
HDA Person Dataset [129]	85	13	Provides tight bounding boxes and occlusion flag	No
Shinpuhkan Dataset [130]	24	16	The tracklets are not followed frames	No
CASIA B [123]	124	11	The scenario includes normal variation in clothing and carrying a bag condition	Yes
MARS [131]	1,261	6	All bounding boxes and tracklets are automatically generated	No

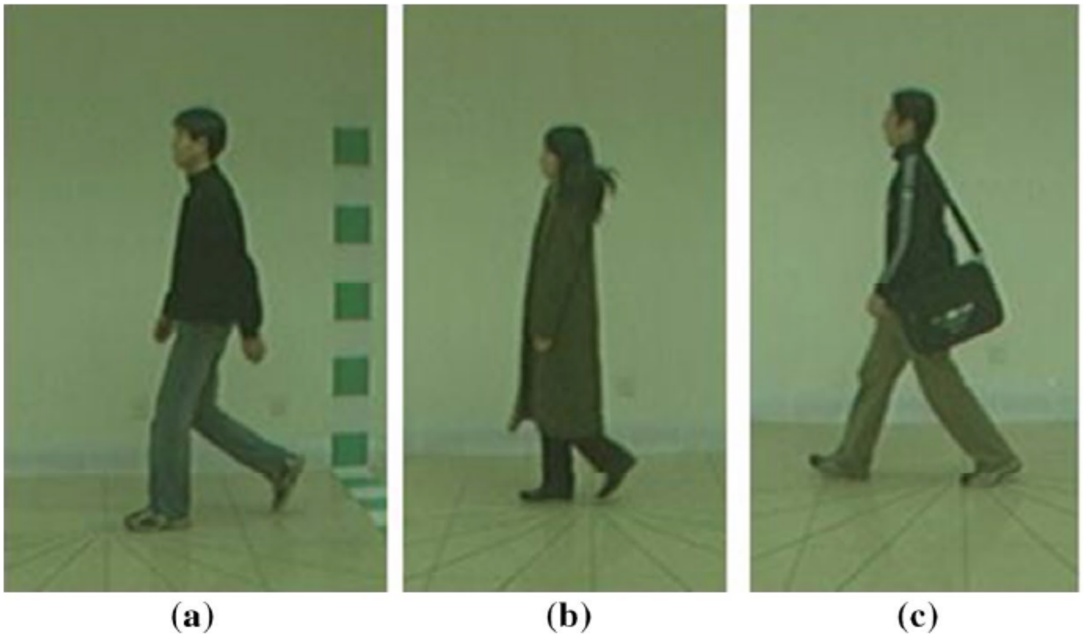


Figure 3.3: (a) normal scenario, (b) heavy clothes scenario, and (c) bag scenario [123]

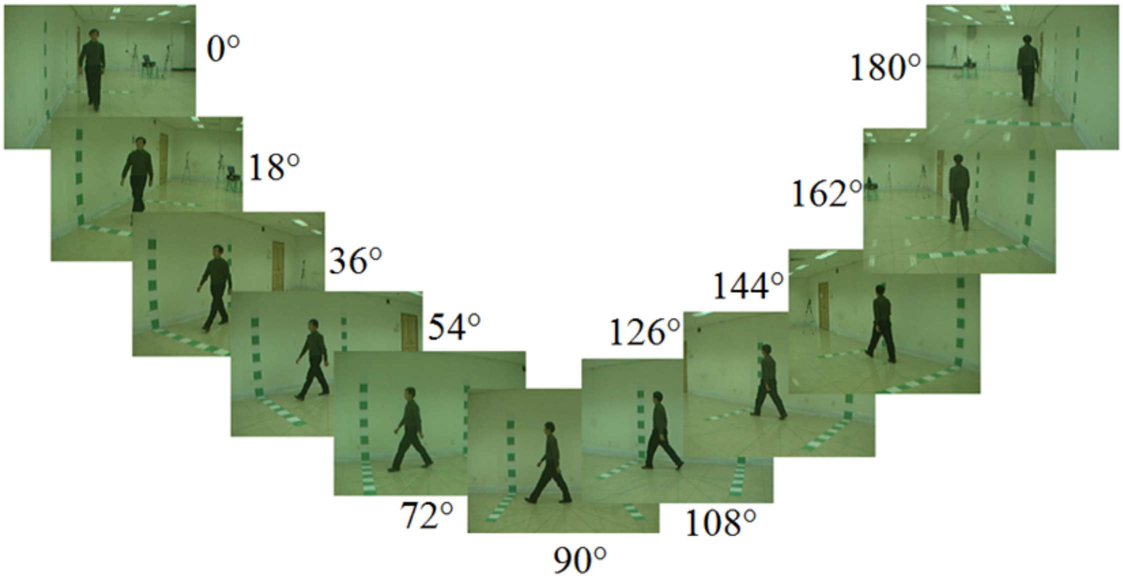


Figure 3.4: CASIA Dataset B viewing angles [123]

3.4 Silhouette Segmentation

The literature on person re-id research illustrated in Chapter 2 shows that there are various ways to build a unique signature for each exiting subject in a scene. Most of the studies reported on signatures developed based on colour, texture, or gait features. However, reviewing the literature on use of body shape features revealed few studies in which shape descriptors on the human body shape were utilised, with [53], [55], and [57] being among them. These studies were mainly focused on identification (i.e., one-to-many process) using mmW images, which have a higher quality than normal CCTV video stream images, while the aim of this research is to re-identify subjects (many-to-many process) in public areas and to use images that simulate the quality of the CCTV images. Therefore, several research questions remain open as a result of this research gap:

- What part/s of the human body (obtained through image segmentation) deliver the most discrimination information?

- Considering the natural movement of people in a public area, viewing angle (the side from which the subject is recorded) is the main factor that might affect the shape of the human body. To what extent it is effective to use the digital description of the shape of these parts for true matching in an environment such as public areas?

Consequently, it was decided that the shape of human body parts (segments) and their movement be investigated for person re-id.

Current systems consider the human body in a silhouette frame as one single, connected region. This assumption increases motivation for body segmentation, mainly because segmenting the human silhouette allows for the independent discovery of the static and dynamic features of different segments of the human body. In the previous chapter, several human body segmentation methods were reviewed. One of the common body segmentation methods was to segment according to the length of the bounding box that surrounds the silhouette. in this research, the length of each segment is extracted from four anthropometry studies to justify the starting end points and length of each segment.

In this research, it is proposed that the human body is divided into four sections that do not overlap, meaning the segments have no common zones. The segments are designed to be non-overlapping to avoid performance duplication and to present a fair comparison between each segment's accuracy.

The human body can be divided into a number of different segments varying in length and size. Although all body parts move accordingly with a human's motion, some parts of the body show motion more than others, such as the amount of the motion in the head and in the leg. Therefore, every part of the body that shows motion is segmented as a separate section for further investigation. Since in this research the accuracy of utilising the digital representation of the shape of the body parts is being examined, the proposed segments are *Head and Neck*, *Shoulders*, *Middle* and *Lower*.

Figure 3.5Figure 3.2 shows the zone of each segment, and the precise length of each segment is explained in Section 3.4.2.

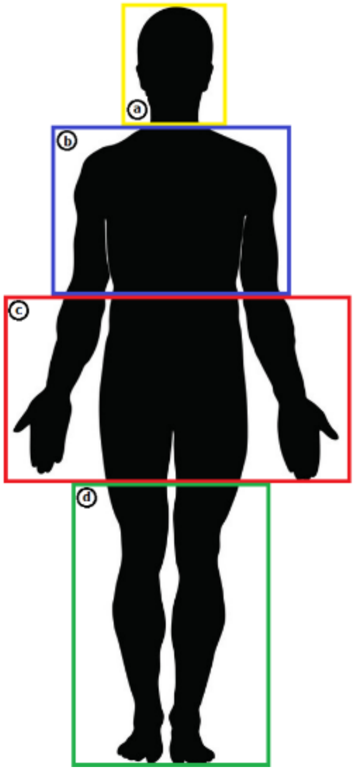


Figure 3.5: Locations of proposed segment zones with respect to the entire body: a) *Head and Neck*, b) *Shoulders*, c) *Middle*, and d) *Lower*

Thus, this section explains the human body shape segmentation used in this research; four anthropometry studies for anatomical segmentation are reviewed, which illustrate how the silhouettes are arithmetically divided into four segments. It also describes what body segments are considered and what the justifications are for each segment.

3.4.1 Body Segments' Length Parameters in Anthropometry

The human body consists of several segments or links connected by joints. Length, weight, and volume are values or parameters that describe these links or segments. Anthropometry is a branch of anthropology that deals with measuring human body parameters. This is mainly used in anthropological categorisations and comparisons between populations.

There are many anthropometrical studies that measure body segments' parameters. However, since the human body is being divided into parts based on the segment length parameter for this research, the studies that provide body part length percentages with respect to the entire body are most appropriate. Therefore, the results from four such studies are averaged and used for silhouette segmentations. A brief overview of these studies, including the number, gender, and ethnic group of the subjects, can be found in the following tables. Table 3.2 shows the anthropometrical studies reviewed, along with the total number of subjects, and provides a breakdown of the examined populations by gender and ethnicity.

Table 3.2: Four anthropometrical studies synthesised and used for silhouette segmentations in this research.

Study #	Study By	Total # Subjects	Ethnicity	# Male	# Female
1	Drillis & Contini [132]	20	American	20	-
2	Dempster [133]	9	White	9	-
3	Contini [134]	29	American	21	8
4	De Leva [135]	115	White	100	15

Table 3.3 summarises the average percentage length of each body segment, considered in this research from four anthropometry studies. The percentage shown in the table is the percentage of the segment with respect to the entire body. These lengths were collected from different ethnic groups as well as from both genders to ensure that the overall average length was generalised. The differences in the percentage length between males and females presented in Table 3.3 illustrate that the differences are insignificant (i.e., less than 1.5% in all studies) when comparing the percentage of the segment to the entire body. Therefore, the male and female percentage length parameters are equally averaged. The overall average percentage lengths for the proposed segments were then used for arithmetically dividing the silhouettes, as explained in the next subsection.

Table 3.3: The percentage of each segment from the four anthropometrical studies

Study #	**1**		**2**		**3**		**4**		**Overall Avg.**
Segment	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	
Head and Neck	15.6	-	15.5	-	15.5	15	16.7	17.9	**16%**
Shoulders	22.2	-	20.3	-	21.5	21.8	16.6	17.1	**20%**
Middle	24.5	-	28.2	-	27.3	26.2	25.7	25.6	**26.25%**
Lower	37.7	-	36	-	35.7	37	41	39.4	**37.75%**

3.4.2 Segments Description

This research explores the effectiveness of the entire body shape in person re-id; it also investigates the feasibility of different parts of the body shape and how factors such as viewpoint affect accuracy rates. Hence, the body silhouette is divided into the proposed four non-overlapping segments, which include *Head and Neck*, *Shoulders*, *Middle*, and *Lower.*

These segments are designed to be non-overlapping to avoid performance duplication and to present a fair comparison between the segments. The precise start and end points and the length are built based on four anthropometry studies.

Table 3.4 presents the proposed segments, their length percentage with respect to the entire body, and the start and end percentage points according to the body. The table also gives a physical description that shows the precise zone in the body and the aimed shape and movement that this segment covers.

Table 3.4: Summary of the proposed body segments

Segment	**% of Total Body**	**Starts%**	**Ends%**	**Seg. Description**	**Seg. Coverage**
Head and Neck	15.9	0	15.9	Starts from the top of the body (i.e., top of the head) and ends at the point that separates the neck from the trunk.	The shape of the head and the neck
Shoulders	20.25	15.9	36.1	Starts from the point that separates the neck from the trunk and ends at the line that connects the elbows.	The shape of the shoulders, the upper arm
Middle	26.3	36.1	62.4	Starts at the line that connects the elbows and ends at the line that connects the fingertips.	The shape of the forearms and hands
Lower	37.6	62.4	100	Starts at the line that connects the fingertips and ends at the bottom of the body (i.e., the sole).	The shape of the feet, lower legs, and parts of the thighs, excluding the hand.

3.4.3 Arithmetic Silhouette Segmentation

The silhouette sequences provided by CASIA Dataset B from 11 different views were used for this research. The silhouettes were algorithmically divided as follows: after detecting the silhouette in the image, the silhouette is cropped by finding the smallest bounding box, where the width (horizontal) and height (vertical) of the bounding box are represented as the x and y axis of the image, respectively. This means that the sizes of the bounding boxes were different, depending on the silhouettes' sizes. Then the silhouettes were divided into the four proposed segments based on the height y of the silhouette, with y here representing the height of the entire body. These segments were chosen to represent the different non-overlapping sections of the human body to enable an examination of which delivered reliable person re-id data. The segments' height measurements were based on the vertical extent of the bounding box. The extracted segments were *Head and Neck*, *Shoulders*, *Middle,* and *Lower*. To automate the segmentation, these segments were defined using arithmetic operations that normally used silhouette for gait recognition techniques [136]. The top segment (i.e., *Head and Neck*) is segmented thus:

$$I(x,y) = \begin{Bmatrix} (x,y)|x_c - \frac{w}{2} \leq x \leq \frac{w}{2}, \\ y_c + \frac{h}{2} \leq y \leq y_c + \frac{h}{2} - h\epsilon_1 \end{Bmatrix} \tag{3.1}$$

where $I(x,y)$ is the bounding box frame with width w and height h. Its centre is (x_c, y_c) and it can be computed thus:

$$(x_c, y_c) = \left(\frac{w}{2}, \frac{h}{2}\right) \tag{3.2}$$

The middle segments (i.e., *Shoulders* and *Middle*) are computed as follows:

$$I(x,y) = \begin{Bmatrix} (x,y)|x_c - \frac{w}{2} \leq x \leq \frac{w}{2}, \\ y_c + \frac{h}{2} - h\epsilon_1 < y \leq y_c + \frac{h}{2} - h(\epsilon_1 + \epsilon_2) \end{Bmatrix} \tag{3.3}$$

The bottom segment (i.e., *Lower*) can be extracted thus:

$$I(x,y) = \begin{Bmatrix} (x,y)|x_c - \frac{w}{2} \leq x \leq \frac{w}{2}, \\ y_c + \frac{h}{2} - h(\epsilon_1 + \epsilon_2) < y \leq y_c - \frac{h}{2} \end{Bmatrix} \tag{3.4}$$

where ϵ_1 and ϵ_2 are constants that can determine the height of each segment in relation to the entire bounding box. These constants are determined using the segments' length parameters from the anthropometry studies, which was discussed in the previous subsection.

3.4.4 Implementation

Figure 3.2 shows the framework of the proposed shape-based person re-id. This Chapter focuses on the first stage of the proposed system, which is silhouette preparation and segmentation. The preparation processes is implemented using a graphical user interface (GUI) in MATLAB. First part involves obtaining the smallest bounding box that surrounds the subject in each frame in the dataset. This is applied using *regionprops* method that receives a silhouette frame containing one subject. Using the property *BoundingBox* with the method *regionprops*, it returns the smallest rectangle containing the subject's silhouette. The segmentation process is presented in sections 3.4.1, 3.4.2, and 3.4.3. In this Section the practical steps of implementing the preparation and segmentation processes are listed as follows:

1. All 11 views provided in the dataset are included for each subject.

2. Normal, wearing heavy clothes, and carrying a bag scenario are included.

3. All the subjects in the dataset are included.

4. Each subject is provided with two different sequences from each angle in each scenario.

5. For every frame, the bounding box, which is the closest box that surrounds the silhouette, is cropped (see Figure 3.6).

6. The proposed silhouette segmentations are implemented as stated in Section 3.4.3 (see Figure 3.7).

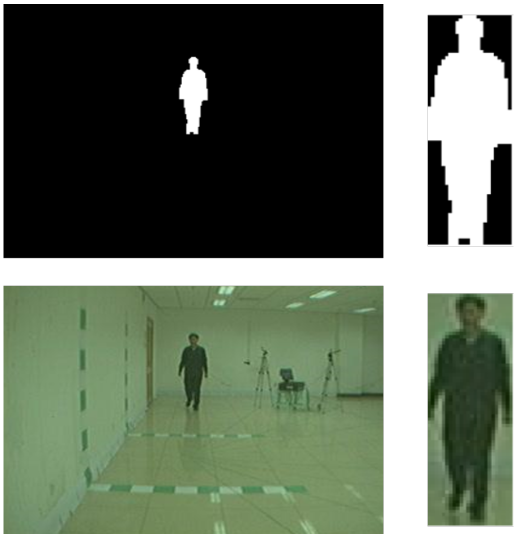


Figure 3.6: Extracting the smallest bounding box that surrounds the subject in each frame

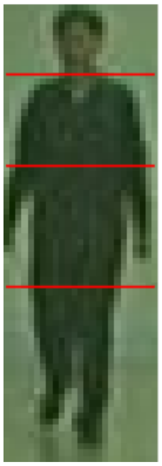


Figure 3.7: Segmentation implementation

3.5 Summary

Investigating the accuracy of the shape descriptor as an identifier for body segments requires a silhouette segmentation that involves two pre-processing procedures. The first procedure involves finding the target dataset that provides (a) multiple viewing angles, (b) tracking sequences, (c) silhouette sequences corresponding to the original sequences, and (d) different appearance scenarios. This led to a review of the publicly available person re-id datasets, which then led to the identification of CASIA Dataset B as the target dataset.

The other pre-processing procedure was the silhouette segmentation that divided the human body into four non-overlapping parts. In order to implement this, four anthropometrical studies were reviewed, which provided a numerical percentage for the length of each segment of the human body with respect to the whole body. Consequently, the proposed segments are *Head and Neck*, *Shoulders*, *Middle,* and *Lower*. Finally, the segmentation was implemented over the silhouettes from CASIA Dataset B using arithmetic operations.

Chapter 4

Body Shape Static Variation Using Generic Fourier Descriptor

4.1 Introduction

The literature review revealed a number of important aspects of person re-identification (re-id) implementation. This is considered a new field relative to other biometrics applications, such as fingerprint and iris verification; as a result, there is a lack of knowledge on soft biometrics extraction and the performance of soft biometrics as identifiers for person re-id. One of the soft biometrics that has received limited attention in this context is the body shape descriptor. To address this gap in the literature, this chapter examines the performance of the shape descriptor of the body and body segments for person re-id. For this analysis, the general performance of the body shape descriptor is investigated. Factors that negatively and/or positively influence accuracy when utilising body shape descriptors in person re-id are explored as part of this examination.

Another objective of this study is to identify the performance and effectiveness of shape descriptors for four proposed body segments, namely *Head & Neck*, *Shoulders*, *Middle,* and *Lower,* from different views (i.e. angles). Consequently, this examination presents the accuracy levels for each segment, with consideration of viewing angle changes.

Segment integration is another focus of this research. Segment integration is the process of joining two or more connected segments of the four originally proposed segments, the purpose of which is to determine the identification performance of using wider parts of the body.

Several possible scenarios exist for capturing body images on CCTV camera networks in public areas. For example, a subject in motion might be recorded from one view at one scene and then from another view in the same or a different scene. There is a persistent requirement to examine the ability of the shape description to extract the uniqueness of the body and body segments shapes under such conditions to simulate a real scenario.

Another possible situation for CCTV in public areas is that a subject is recorded in one scene walking normally and then recorded in another scene under different conditions, such as wearing heavy clothes or carrying a bag. Since conditions such as these affect body shape, these scenarios are considered in this research as well.

This chapter assesses the feasibility of the shape descriptor of the body segments in person re-id applications through image-based approach. This approach apply the classification process in frame-by-frame basis, examining the static shape variation within one frame.

This chapter is organised as follows: Section 4.2 details the methodology followed in this research, utilising the Generic Fourier Descriptor and the Linear Discriminant Analysis as classifier. Section 4.3 presents the general performance of shape description utilising the body shape as an identifier for person re-id. Also included in this Section, the GFD descriptor is extracted from the four proposed segments, which are classified to show their performance stability. Section 4.3.3 presents the new segments added to the original set of proposed segments and investigates their accuracy rates. Section 4.4 demonstrates the inter-view performance, where the subject is recorded from two different views. Finally, Section 4.5 implements two scenarios, including carrying a bag and wearing heavy clothes, illustrating the segments most and least affected by changing the scenario.

4.2 Methodology

The main objective of this research is to discover the performance of the shape descriptor as a soft biometric for person re-id. Therefore, the whole body and individual body segment shape descriptors are extracted separately. The four body segments proposed in this research are *Head & Neck*, *Shoulders*, *Middle,* and *Lower*. Body segmentation is explained and justified in Chapter 3.

The shape descriptor used in this experiment is the Generic Fourier Descriptor (GFD) [102]. Shape descriptors in general are divided into contour-based shape descriptors and region-based shape descriptors. Contour-based descriptors only extract information located at the boundary; the interior content of the shape cannot be used. Therefore, contour-based descriptor applications are limited. The region-based shape descriptors are extracted utilising all pixel information in the contour and inside a shape region. Region-based shape descriptors are implemented in more applications than contour-based descriptors. The performance of the GFD was compared in [102] with the common contour-based and region-based shape descriptors. The results showed that GFD outperformed the other two techniques.

The dataset used in this research is CASIA Dataset B [123], discussed in detail in Chapter 3, which involves three scenarios, namely, walking in normal clothing, carrying a bag, and wearing heavy clothes. This dataset recorded subjects from 11 viewing angles and provided two different video records for each subject from each angle. The two videos were recorded under the same conditions but at different times. Having two different image sequences (videos) for the same subject under the same conditions enables the use of one of the sequences for training and the other for the test in the classification process. A Linear Discriminant Analysis (LDA) classifier [104] was applied to each GFD feature vector of an image segment individually. A separate GFD feature vector was extracted for each frame. After training the LDA on the input set (one of the sequences), the error rate was computed by testing the input set against the watch list (the other sequence). The results are shown using the CMC curve explained in Chapter 3.

4.2.1 Generic Fourier Descriptor Implementation

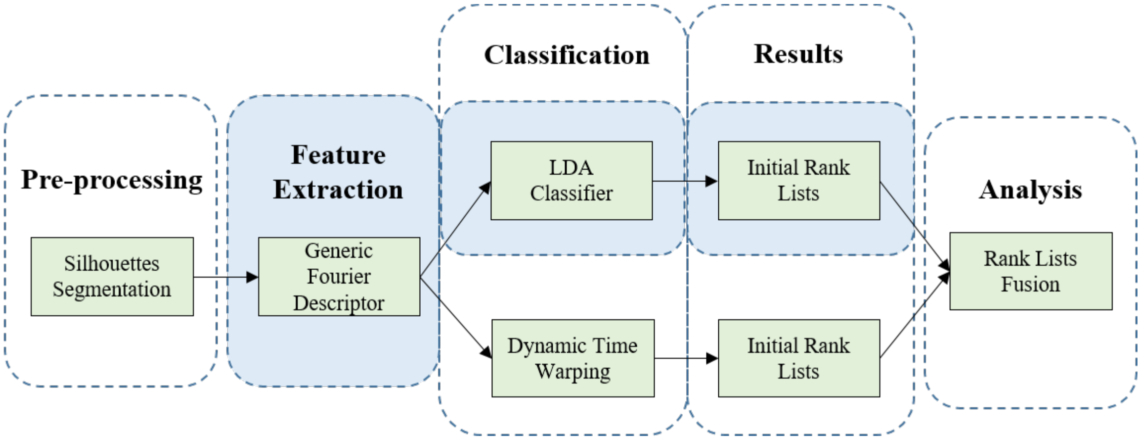


Figure 4.1: Proposed shape-based person re-id system framework

Figure 4.1 illustrates the experimental framework implemented in this research. After obtaining the proposed segments, the feature extraction step was implemented. In this chapter, the feature extracted from each image segment was the GFD. This shape descriptor extracts the properties on the contour and the region of the silhouette. It also generates a fixed length features vector, regardless of the silhouette size. The GFD technique is achieved using modified polar Fourier transform (MPFT) on the silhouette image. To implement MPFT, the original image in Cartesian space is converted into a rectangular image in polar space. The polar coordinates (r, θ) can be extracted from Cartesian coordinates (x, y) as:

$$r = \sqrt{(x - g_x)^2 + \left(y - g_y\right)^2} \tag{4.1}$$

$\left(g_x, g_y\right)$ is the centroid of the foreground image.

$$\theta_i = i(2\pi/T) \tag{4.2}$$

where $0 \leq r < R$.

A 2D Discrete Fourier transform (DFT) is then applied on the polar images to extract Fourier coefficients. DFT are used to create the feature vector that represents the shape.

$$pf(\rho, \varphi) = \sum_r \sum_i f(r, \theta_i) exp\left[j2\pi\left(\frac{r}{R}\ \rho\ +\ \frac{2\pi i}{T}\varphi\right)\right] \tag{4.3}$$

where R is the radial resolution and T is the angular frequency. Then, for each image segment, the generated feature vector is represented as follows:

$$GFD = \left(\frac{pf(0,1)}{pf(0,0)}, \dots, \frac{pf(0,n)}{pf(0,0)}, \dots, \frac{pf(m,0)}{pf(0,0)}, \dots, \frac{pf(m,n)}{pf(0,0)}\right) \tag{4.4}$$

where m and n are the maximum numbers selected of radial resolution and angular frequencies, respectively. By setting these two variables, we control the length of the generated feature vectors to be all of the same length.

Table 4.1: Generic Fourier Descriptor Algorithm

Generic Fourier Descriptor Algorithm

```
Input: BW, X by Y binary image containing single object
Input: m, denotes the radial frequency, where $m > 0$
Input: n, denotes the angular frequency, where $n > 0$
Output: Vector FD is $(m*n+n+1)$

Let BW be logical
Let the object in BW be centered
Let FD be a zeros matrix of size $((m+1)*(n+1),1)$
Let FI be a zeros matrix of size $(m+1,n+1)$
Let FR be a zeros matrix of size $(m+1,n+1)$

$i = 1$
% loop over all radial frequencies
for rad = 0  to m do
    %loop over all angular frequencies
    for ang = 0 to n do
```

```
        % calculate FR and FI for rad and ang
        tempR = BW*cos(2*pi*rad*radius+ang*theta)
        tempI = -1*BW*sin(2*pi*rad*radius+ang*theta)
        FR(rad+1,ang+1) = sum(tempR(:));
        FI(rad+1,ang+1) = sum(tempI(:));
        %calculate FD, where FD(end)=FD(0,0) --> rad == 0 & ang == 0
        if  rad == 0 && ang == 0
           % normalized by circle area
           FD(i) = sqrt((FR(1,1)^2+FR(1,1)^2))/(pi*m^2);

        else
        % normalized by |FD(0,0)|
        FD(i)                                                            =
sqrt(((FR(rad+1,ang+1).^2+FI(rad+1,ang+1).^2)))/sqrt((FR(1,1)^2+FR(1,1)^2));

    end for
    i = i + 1
end for
Return FD
```

4.2.2 GFD resolution versus performance

The number of the radial features and angular frequency are parameters that can determine the resolution of the GFD. In [102], the GFD shape descriptor was used for shape retrieval where the examined shapes are from different classes, such as leaves and animals; the researchers found that a small number of radial and angular features was sufficient to achieve a high level of retrieval accuracy. However, in this research, the GFD descriptor is used on human body and body segments, where all the shapes are within the same class which is human. Therefore, it is significant to select optimal values for the number of the radial and angular features for such shape. A range of numbers of radial and angular features were evaluated on the body shape of all subjects in the dataset; the range of the evaluated radial number was 1 to 10 features, and the range of the angular number was 1 to 35 features. Figure 4.2 shows the relationship between the accuracy rate of the body GFD shape descriptor and the GFD resolution (i.e. the number of the radial and angular features'). For simplicity, the radial number presented in this figure are from 3 to 6 features and the angular number presented are from 10 to 28 features, where the highest accuracy levels were located. This figure shows the As depicted, the accuracy is achieved by employing various feature numbers using equation

$$pf(\rho, \varphi) = \sum_r \sum_i f(r, \theta_i) e\,xp\left[j2\pi\left(\frac{r}{R}\;\rho\;+\;\frac{2\pi i}{T}\varphi\right)\right] \qquad (4.3)$$

The accuracy rate is approximately 70%, even with a small number of features. However, the accuracy rate is more influenced by the number of the angular value (*T*) than by the number of the radial value (*R*). Performance is enhanced by increasing the number of the angular (*T*) to a certain level; then the accuracy rate is suddenly decreased. Consequently, the results reflect that the minimum optimal angular and radial values achieving an effective re-id performance are the value 5 for the radial feature and the value 25 for the angular feature. This produces a GFD feature vector of 156 elements length for each image segment processed in this experiment.

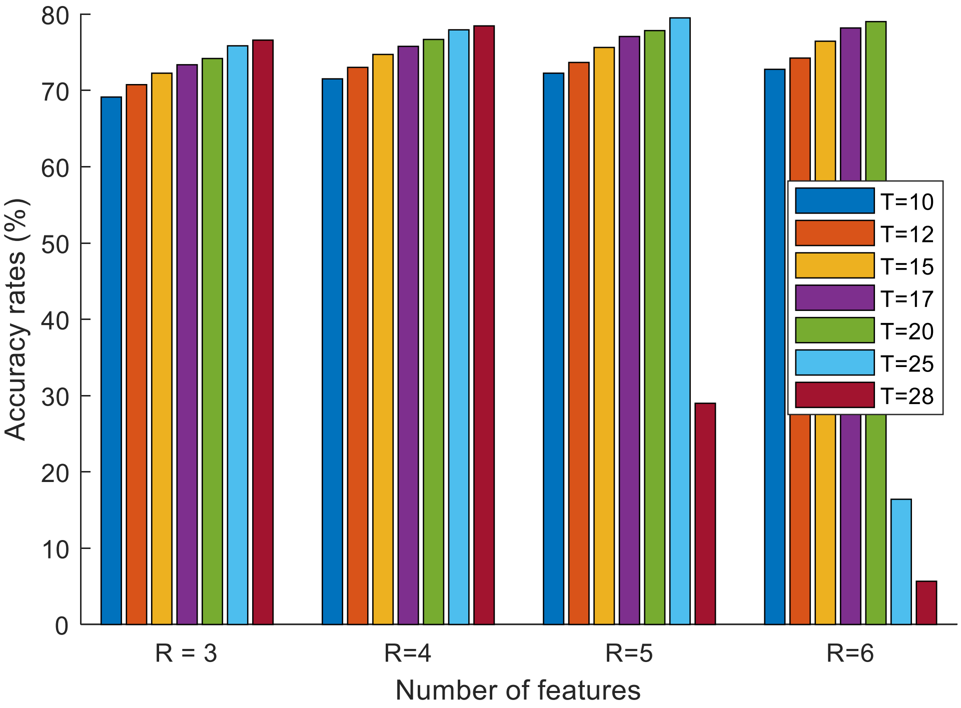


Figure 4.2: Number of features versus performance where R refers to the number of radial features and *T* refers to the number of angular features

4.2.3 GFD Performance Evaluation

A Linear Discriminant Analysis (LDA) classifier [104] was applied to each GFD vector of an image segment individually. LDA mainly maximises the distances between the mean of the classes and minimises the variation within one class's samples.

A separate GFD feature vector was extracted for each frame. As above-mentioned, the CASIA Dataset B provided two sequences for each person that was recorded under the same conditions. The main reason of providing two sequences per person is to train the system on one sequence from each subject in the dataset and test the system based on the other sequence. This was implemented in this work; LDA classifier trained on the training set (i.e. the set containing the first sequence of each subject), and the error rate was then computed by testing the training set against the test set (i.e. the set containing the second sequence of each subject).

The second step of evaluating the system is to compute the ranks, which was don as follows: 1) a matrix of $m * n$ was obtained, where m equalled the number of samples in the training set, and n equalled the number of GFD samples in the test set; 2) this matrix contained the similarity scores of each sample in the test set against each sample in the training set; and 3) the values (scores) in each column of the matrix were then sorted in descending order; 4) replace the scores with their corresponding identities.

4.3 Body and Body Segments in Image-based System

4.3.1 Body Performance

The objective of the first question in this research is to discover the performance of the GFD shape descriptor utilised on a sequence of body silhouettes and to compare its accuracy with state-of-the-art person re-id using soft biometrics as an identifier. Therefore, the GFD of the entire body shape (i.e. silhouette) was extracted from the image sequences provided by CASIA Dataset B. This dataset is discussed in Chapter 3. The image sequences recorded subjects from 11 different views (i.e. angles). The normal scenario was used at this stage, where the subjects were asked to walk normally from one side to another twice. This provides two different sequence for the same subject under the same conditions allowing to use one sequence for test and the other one for training. The GFD shape descriptor corresponding to each segment frame were classified using LDA classifier. The classification process done through image-based approach which classify in frame-by-frame basis, examining the static shape variation within one frame. This, in fact, generates one rank list for each segment frame.

In order to evaluate the performance, Cumulative Matching Characteristic (CMC) curves were used. CMC curves accumulated the numbers of the true re-id for each rank and displayed them in order. Figure 4.3 shows the performance of GFD employed on the body silhouettes from 11 different views. Accordingly, the number of observations can be observed from the results in this figure.

First, the accuracy when using the shape description for person re-id from all views at the first rank was 37-76% and reached the boundaries between 87-97% at rank 20. Second, regarding the individual view performance, the accuracy in the 0° and 180° views were the highest, compared with the other views. According to 0° and 180° view sequences, the subjects were recorded from a straight front or a straight back respectively, for the whole sequence. This means that the body orientation in these two views (i.e. 0° and 180°) remains the same within the sequence, while the rest of the views contained variation in body orientation within the same sequence. Therefore, it can be observed that the accuracy rates for the rest of the views were less than those in the 0° and 180° views. It can be concluded, then, that body orientation is one factor that affects the accuracy rate when using the body shape description as an identifier in person re-id.

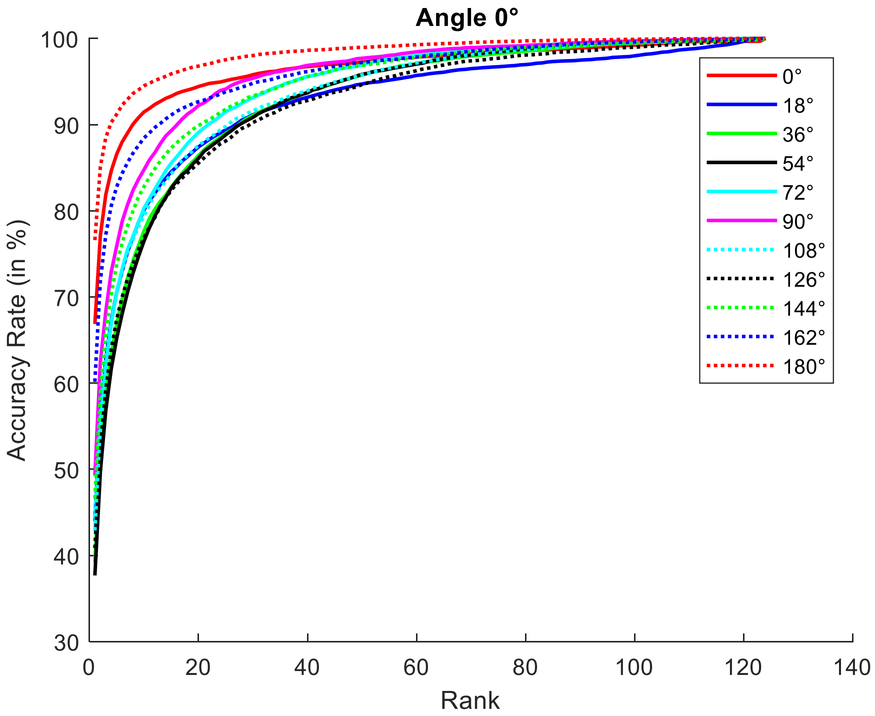


Figure 4.3: CMC curves showing the accuracy of the GFD descriptor used on the body silhouettes from 11 different viewing angles

4.3.2 Segments Performance

Extracting the GFD shape descriptor from the body silhouette for use as a soft biometrics identifier shows comparable results to state-of-the-art person re-id. This indicates a need to investigate the performance of different segments (or parts) of the body and whether the accuracy varies between the segments. Therefore, the four body segments (*Head & Neck*, *Shoulders*, *Middle,* and *Lower*) proposed and explained in Chapter 3 are employed in this research to answer this question.

Each segment is processed as a separate silhouette frame, where the GFD descriptor is extracted and classified using the LDA classifier, following the whole-body shape methodology. The accuracy rate of each segment is individually presented using CMC curves.

The results are depicted in four segments that form 11 views to display. All results of the first rank accuracy rate from each segment are shown in Table 4.3. To reduce the ambiguity, only four views are visually presented, including 0°, 54°, 90°, and 162°, where the main observations lie. The rest of the views delivered similar outcomes to the presented views.

Figure 4.4 - Figure 4.7 demonstrate the accuracy rates of the *Body, Head & Neck, Shoulders, Middle,* and *Lower* segments from 0°, 54°, 90°, and 162° views respectively.

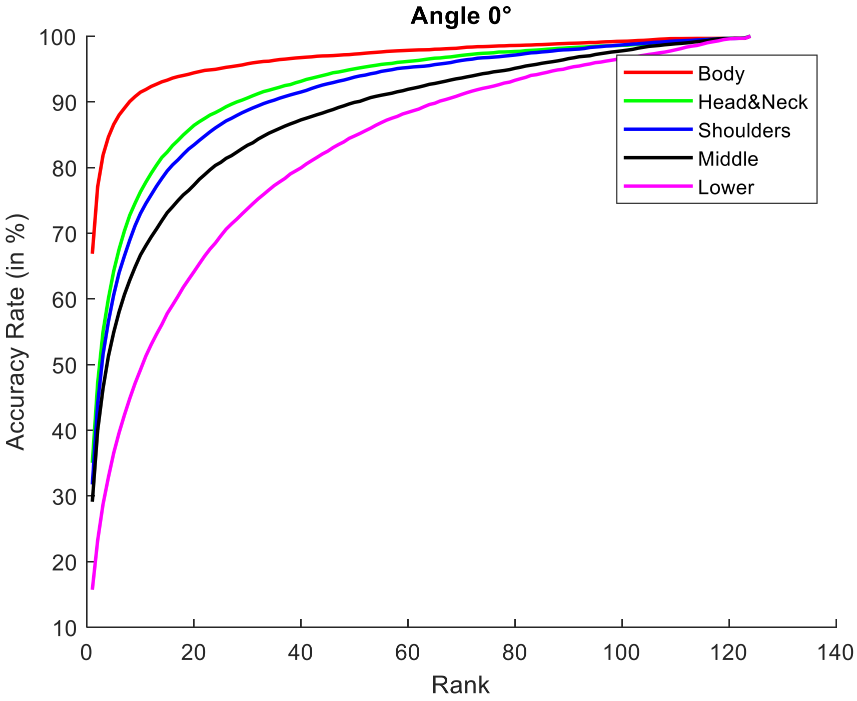


Figure 4.4: The performance of the *Body*, *Head & Neck*, *Shoulders,* and *Middle* segments from the view 0°

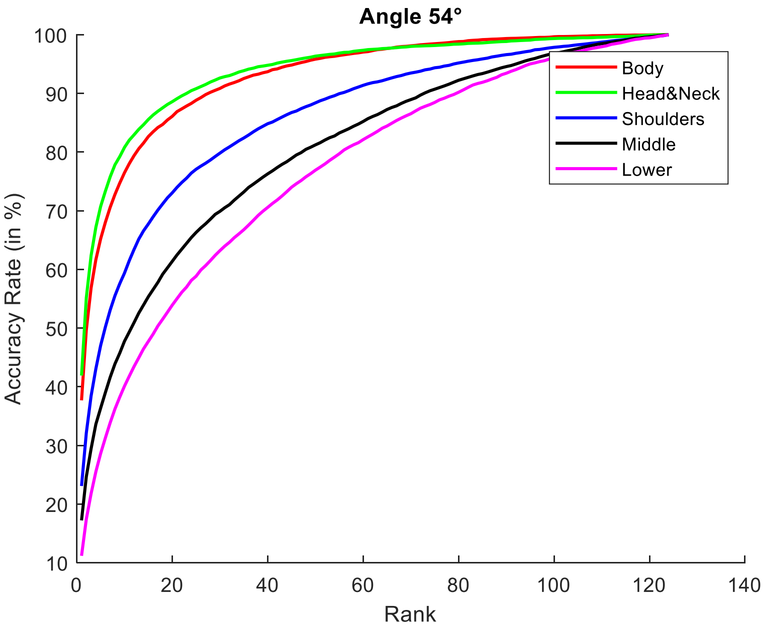


Figure 4.5: The performance of the *Body*, *Head & Neck*, *Shoulders,* and *Middle* segments from the view 54°

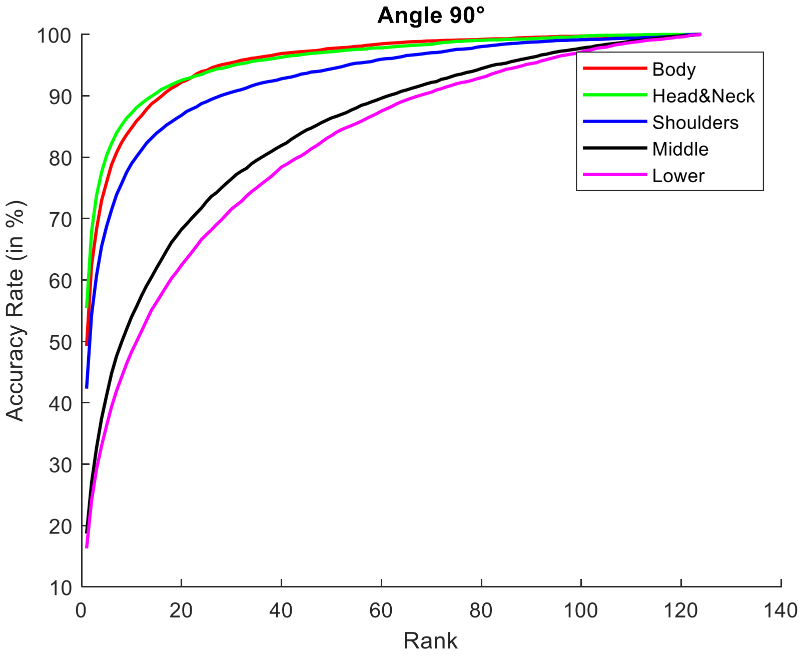


Figure 4.6: The performance of the *Body*, *Head & Neck*, *Shoulders,* and *Middle* segments from the view 90°

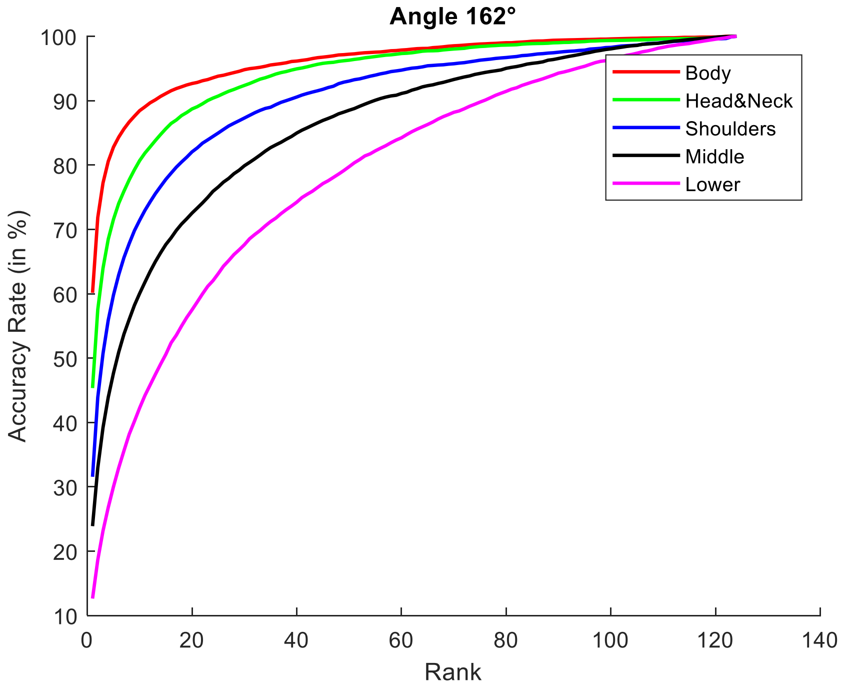


Figure 4.7: The performance of the *Body*, *Head & Neck*, *Shoulders,* and *Middle* segments from the view 162°

Based on the curves shown in

Figure 4.4 - Figure 4.7, a number of findings can be observed. However, the main observation is the variation in the distinction levels in between the segments, as different segments within the same body displayed different accuracy rates. In the following subsections, each segment is analysed in terms of its stability, performance accuracy, and main influences. All results of the first rank accuracy rate from each segment are shown in Table 4.3.

Body; Compared to the other segments, the *Body* segment outperformed in the straight views (i.e. 0° and 180°) when the body orientation does not change within the sequence. In the straight views, the *Body* segment achieved 67% and 76% accuracy rates, respectively. Also, the *Body* shape accuracy rates decreased in the side view, primarily because the hand and leg shape, when the subject is in motion, have more influence on the side view than on the straight views. Thus, the *Body* segment produced unstable levels of accuracy influenced by body orientation and the hand and leg shape. Yet high accuracy rate comparing to other segments.

Head & Neck; In general, the *Head & Neck* segment maintained more stable levels of accuracy in all views than the *Body* segment. This means that body orientation has less impact on this segment than on the *Body* segment. Consequently, the *Head & Neck* segment outperformed the *Body* segment in the view where body orientation negatively affected *Body* segment performance (i.e. 54°, 72°, 90°, 108°, and 126°).

Shoulders; The *Shoulders* segment also exhibited performance stability between the views with slight improvements in the side views, especially at 90°. Therefore, the *Shoulders* segment was not influenced by body orientation, either. However, the *Body* and *Head & Neck* segments performed better than the *Shoulders* segment in all views.

Middle; The *Middle* segment is the second least discriminative of the segments because of the low level of accuracy rates. The *Middle* segment is affected by the shape variation caused by hand motion. However, it displayed a performance improvement in the straight views (i.e. 0° and 180°), which means that body orientation is another factor that influences the *Middle* segment.

Lower; Compared to other segments, the *Lower* segment consistently showed the lowest accuracy rates, achieving around 10% in all views, because this segment contains the highest shape variation in the body due to leg movement while the subject in motion. Mainly, this experiment extracted the GFD shape descriptor of the segments and processed the classification using the LDA classifier, which means that this study examined the static variation of the shape between one frame and another. Therefore, the *Lower* segment reflected the lowest level of discrimination and accuracy rates in all angles.

4.3.3 Segments Integration

In the preview section, the performance of the individual body segments' shape description was discovered, and the results indicated that different segments within the same body have different accuracy rates. For example, the *Head & Neck* and *Shoulders* segments consistently performed better than the *Middle* and *Lower* segments in all views. This finding prompted the integration of two or more connected segments from the proposed segments to produce new segments. Two important reasons justify introducing the new segments at this point in the research. First, the new proposed segments were not presented as separate segments in the anthropometrics studies reviewed in this research. They are proposed here as a fusion based on the number of anthropometrical segments. Second, the new segments consist of two or more connected parts of the body, which means that the new segments overlap with the original segments. For a fair comparison, the non-overlapped segments (i.e. the original proposed segments) were presented and examined first.

The newly proposed segments consist of two or three connected segments, which can be used to examine the performance of wider parts of the body and to demonstrate the performance of the outperforming segments integration. The four new segments proposed are Upper quarter *(Upper Q)*, Upper half *(Upper H)*, *Torso,* and Lower half *(Lower H),* see Table 4.2. The body parts encompassed in each segment are shown in Figure 4.8 - Figure 4.11 display the accuracy rates of the new segments along with the original segments from views 0°, 54°, 90°, and 162°. The first rank of the accuracy in each segment from all views is stated in Table 4.3, as well as the start and end of each.

The curves presented in the views 0°, 54°, 90°, and 162° demonstrate the number of findings, mostly reflecting improvements in many aspects. In the following section, each segment is analysed in terms of its stability, performance accuracy, and main influences. All results of the first rank accuracy rate from each segment are shown in Table 4.3.

Technically, these segments follow the same arithmetic operation (explained in Chapter 3) used to segment the originally proposed segments. The GFD shape descriptor was individually extracted from the new segments in the same manner as it was from the original segments. The classification process using the LDA classifier was also implemented on the training and test sets of the GFD feature vectors of each segment, where different image sequences were assigned to the training and test sets. The performance of the new segments is presented using CMC curves and the generated rank lists.

Table 4.2: Description of the new proposed segments

New Seg. Name	Consists of	Starts	Ends
Upper Quarter (*Upper Q*)	Upper part of the body **without** lower arms and hands, including *Head & Neck* and *Shoulders* segments	*Head & Neck*	*Shoulders*
Upper Half (*Upper H*)	Upper part of the body **with** lower arms and hands, including *Head & Neck, Shoulders,* and *Middle* segments	*Head & Neck*	*Middle*
Torso	Upper and lower arms, hands, and trunk, including *Shoulders* and *Middle* segments	*Shoulders*	*Middle*
Lower Half (*Lower H)*	Lower part of the body, including the lower arms and the legs, including *Middle* and *Lower* segments	*Middle*	*Lower*

Upper Q; this segment includes the *Head & Neck* and *Shoulders* segments, covering the shape of the upper part of the body, excluding the hands. In general, the *Upper Q* segment demonstrated a stable level of accuracy in all views, as the *Head & Neck* and *Shoulders* segments showed in section 4.3.2. The *Upper Q* segment outperformed the *Body* segment in all views except the straight views.

Upper H; this segment includes the *Head & Neck, Shoulders,* and *Middle* segments, covering the shape of the upper part of the body, including the hands. The *Upper H* segment indicated the best discrimination levels compared with the other segments, with results very close to the side views of the *Upper Q* segment. Comparing the results of the *Upper Q* and *Upper H* segments reveals that adding the lower arms and hands improves the performance by 6% on average.

Torso; the *Torso* segment consists of the *Shoulders* and *Middle* segments. This segment covers the performance of the integration of the shoulders, upper arms, lower arms, and hands. Although the *Torso* segment resulted in lower accuracy rates than the top accurate segments (i.e. *Upper Q* and *Upper H* segments), the *Torso* segment demonstrated performance improvements to the *Shoulders* and *Middle* segments as its encompassed segments.

Lower H; Although the *Lower* segment presented very low accuracy rates in comparison with all segments, adding the lower arms and hand slightly improved the performance. While both the *Lower* and *Middle* segments contained considerable shape variation, integrating them improved *Lower H* discrimination. This improvement may be due to the increase in the covered area, which increases the chance of extracting accurate features as well.

To summurise, experimenting the performance of different segments of the same body showed variations of the performance of body segments. In addition to the entire body shape descriptor, the *Head & Neck* segment provides an acceptable level of discrimination, outperforming *Shoulders*, *Middle* and *Lower* segments in the side views. These results encouraged the integration of different segmentations, where two or three connected segments are joined together thus producing four more segments, which are *Upper H, Upper Q, Torso and Lower Half*. The performance of the added segments indicates a significant improvement, even in the segments that contained high shape variations, such as the *Lower Half* and *Torso* segments, caused by the subject's motion. In general, segments performed better in the straight views than the side views, mostly due to body orientation within the same sequence.

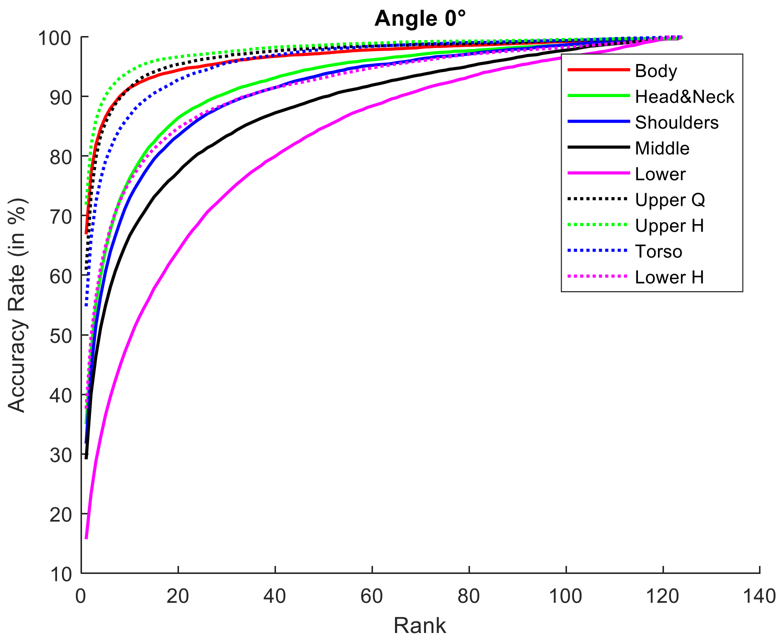


Figure 4.8: CMC curves of performance of the new segments along with original segments from the view 0°

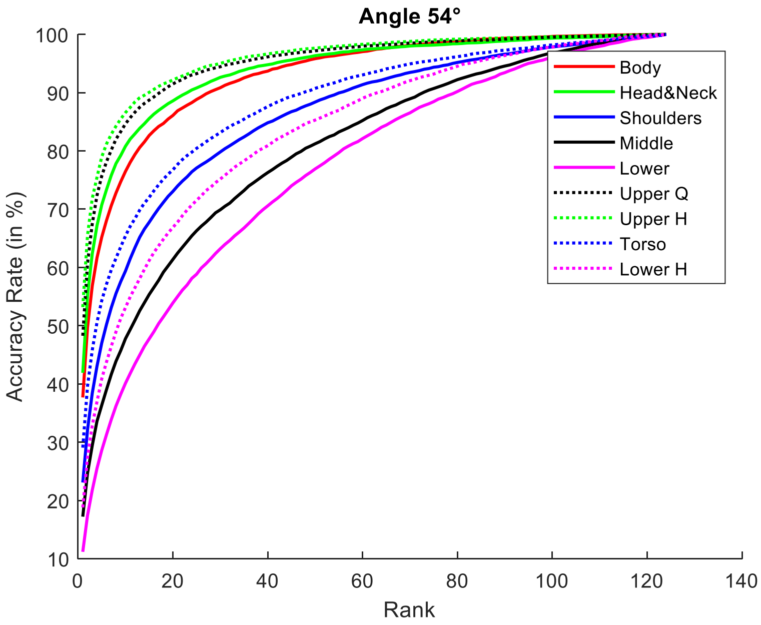


Figure 4.9: CMC curves of performance of the new segments along with original segments from the view 54°

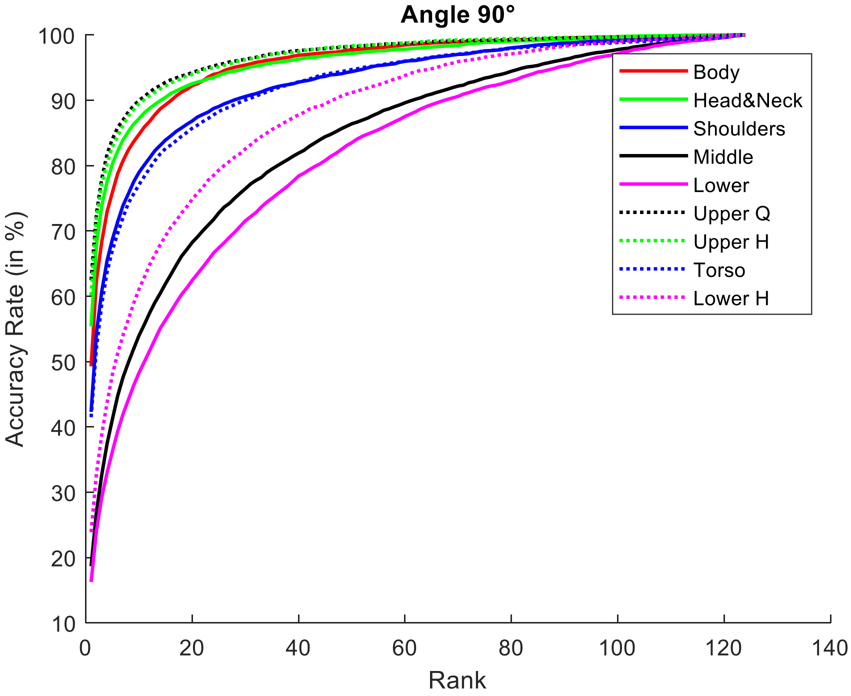


Figure 4.10: CMC curves of performance of the new segments along with original segments from the view 90°

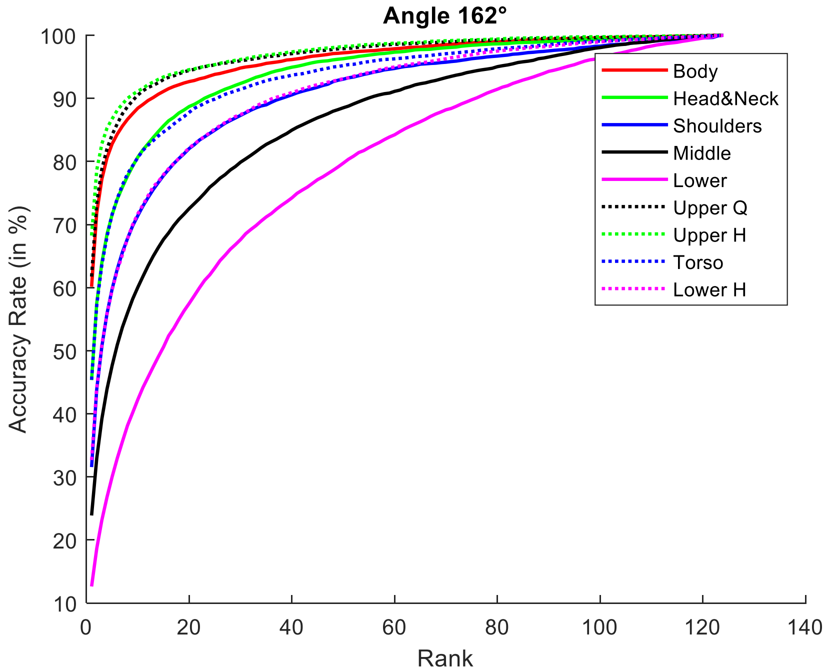


Figure 4.11: CMC curves of performance of the new segments along with original segments from the view 90°

Table 4.3: The first rank of the accuracy rates (in percentage) from all the views (angles)

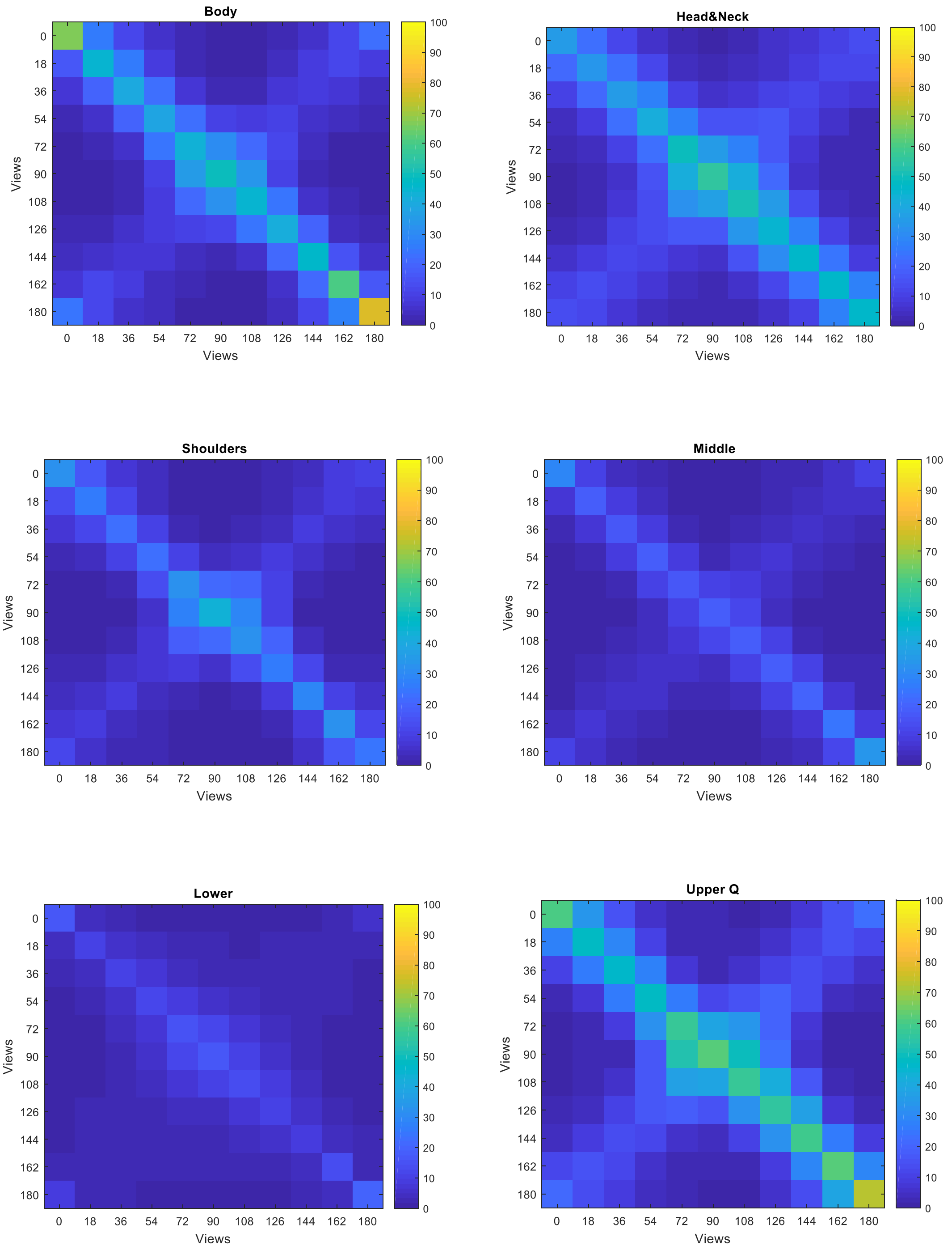
Segment **Angle**	**Body**	**Head & Neck**	**Shoulders**	**Middle**	**Lower**	**Upper Q**	**Upper H**	**Torso**	**Lower H**
0°	0.67	0.35	0.32	0.29	0.16	0.61	**0.72**	0.50	0.38
18°	0.44	0.34	0.26	0.18	0.10	0.48	**0.54**	0.36	0.24
36°	0.39	0.36	0.23	0.16	0.1	0.47	**0.53**	0.32	0.20
54°	0.38	0.42	0.24	0.17	0.11	0.48	**0.53**	0.29	0.19
72°	0.43	0.49	0.32	0.17	0.15	**0.57**	0.56	0.35	0.23
90°	0.49	0.55	0.42	0.19	0.16	**0.62**	0.60	0.42	0.24
108°	0.44	0.51	0.33	0.18	0.13	**0.57**	0.56	0.35	0.20
126°	0.40	0.44	0.25	0.17	0.10	0.55	**0.57**	0.34	0.18
144°	0.46	0.46	0.28	0.19	0.09	0.58	**0.62**	0.38	0.22
162°	0.60	0.45	0.31	0.23	0.12	0.61	**0.68**	0.45	0.32
180°	0.76	0.45	0.24	0.34	0.20	0.73	**0.79**	0.55	0.48

4.4 Inter-views Performance

In the previous section, the general performance of the body shape descriptor indicated comparable accuracy rates for the state-of-the-art person re-id using soft biometrics as an identifier. In addition, it was proven that the shape description of the proposed segments taken from the same body have different levels of discrimination. Consequently, four new segments were added to the original segments, demonstrating wider parts of the body and presenting the performance of the integration of the outperforming segments.

So far, all performances illustrated are considered view-based classifications. This means that the subject image sequences assigned to the training and test sets at the classification stage were taken from the same view. Accordingly, to reveal the answer to the next research question, the capability of GFD shape descriptors of the body and body segments when the subject is recorded from different views was investigated. This is, in fact, a possible scenario for CCTV camera networks in public areas, where a subject in transition might be recorded from one view in one scene and then from another view in the same or a different scene. In order to simulate realistic scenarios, there is a persistent requirement to examine the ability of the shape description under such conditions towards shape based person re-id.

In order to implement the inter-view classification, all the 11 views were tested against each other, the 11 different views can be found in Figure 3.4. The GFD descriptor of each segment was calculated from all views, in the same way as discussed in Sections 4.3.1, 4.3.2, and 4.3.3. Within a segment, each view was tested against all views. This means that if the training set was assigned to a certain view, the test set was assigned to a different view. Each view was tested against the rest of the views for the same body segment. The results are shown based on each segment, for the purpose of identifying how each segment handled the subject view change. Figure 4.12 presents the inter-view segment-based classification as a coloured map. In this figure, the classification of each segment of the originally proposed segments and the new added segments are individually shown.


Body
Head&Neck
Shoulders
Middle
Lower
Upper Q
Views
0
18
36
54
72
90
108
126
144
162
180
0
10
20
30
40
50
60
70
80
90
100

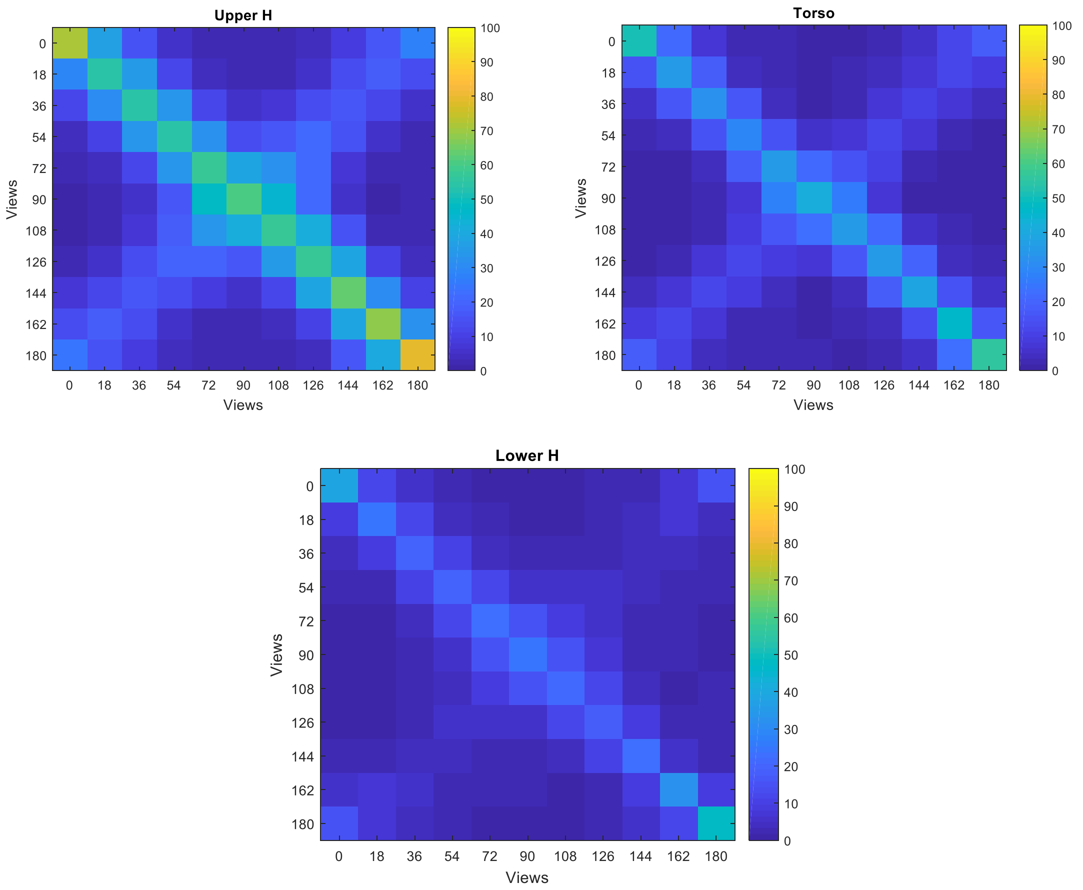


Figure 4.12: Inter-view performance of shape description of the segments

First, the coloured maps depicted in Figure 4.12 demonstrate the general performance of each segment. In general, the coloured maps of all segments reflect a number of findings. Most importantly, the more the tested view is close to the training view, the more the performance is improved, and vice versa. This is clearly illustrated in the coloured maps: the diagonal cells present the view-based classification, where the highest performance occurred (i.e. the training and test sets were from the same view). The cells around the diagonal reflect the performance of the current view against the previous and subsequent views; for example, the views around 90° are 72° and 108°.

Regarding the individual segment performance, the *Upper H* and *Upper Q* segments achieved higher accuracy rates; comparing them with their encompassed segments (i.e., *Head & Neck*, *Shoulders,* and *Middle*) indicates clear accuracy enhancement. However, comparing the results of the *Lower* and *Middle* segments with their integration in the *Lower H* segment signifies that the

Lower H segment slightly outperformed the *Lower* segment but not the *Middle* segment. Similarly, the *Torso* segment did not indicate that performance improved from its original segments—the *Shoulders* and *Middle* segments.

4.5 Cross-Scenarios

In public area CCTVs, one possible situation is that a subject is recorded in one scene walking normally and then in another scene under different conditions, such as wearing heavy clothes or carrying a bag. As wearing different clothes and carrying a bag affect the body shape, these scenarios are considered in this research.

CASIA Dataset B provided three scenarios, namely, Normal, Bag, and Clothes scenarios. In the Normal scenario, the subjects were simultaneously recorded from 11 views while they walked normally. This scenario is the one utilised in all the preview experiments. The Bag scenario is recorded under the same environmental conditions of the Normal scenario; however, the subjects were asked to carry either a handbag or backpack on their shoulder(s). The Clothes scenario was recorded in the same way as the Normal scenario, but the subjects were asked to wear heavy clothes (coats) while they were walking. The Normal, Bag and Clothes scenarios were shown Figure 3.3.

This section aims to discover the ability of the shape description of the body and body segments to uniquely identify the same subject when recorded carrying a bag or wearing different (heavy) clothes. In the following section, the results of the Normal scenario tested against the Bag scenario and the Clothes scenario are provided.

4.5.1 Normal vs. Bag

Previously, the LDA classifier was trained and tested on Normal scenario sequences. In this section, the classifier was trained on one sequence of the Normal scenario and tested on one sequence of the Bag scenario, both are from the same viewing angle. Figure 4.13 - Figure 4.16 show the CMC curves of all segments from four views, including 0°, 54°, 90°, and 162° respectively. Also, Table 4.4 provides the first rank of the accuracy rate of each segment in all views.

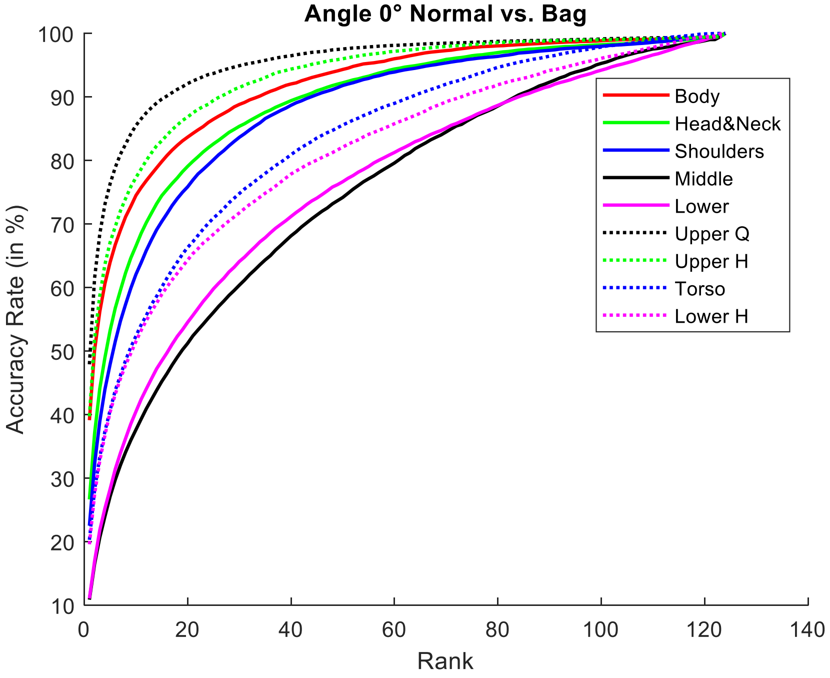


Figure 4.13: CMC curves of the cross-scenario classification showing the results of the Normal Scenario against the Bag Scenario from 0° view

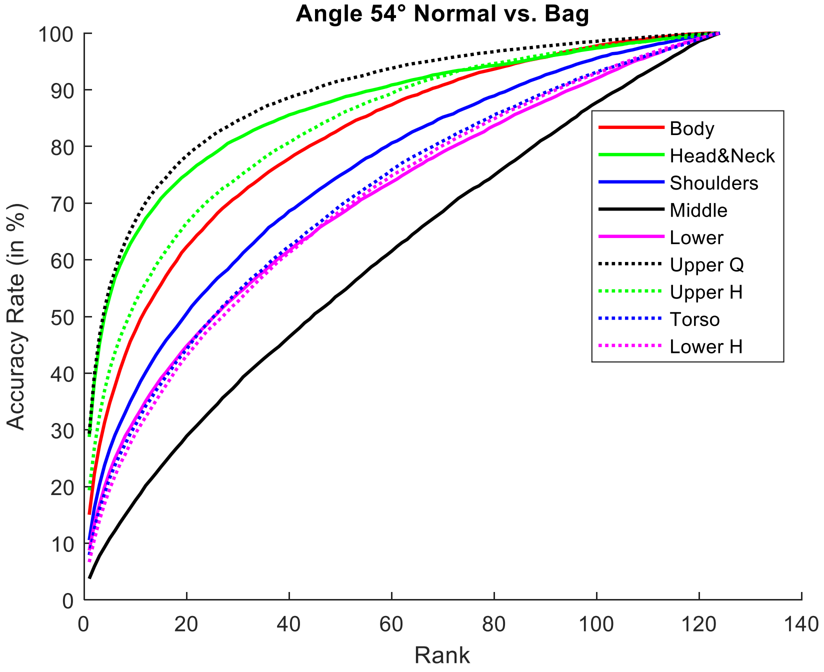


Figure 4.14: CMC curves of the cross-scenario classification showing the results of the Normal Scenario against the Bag Scenario from 54° view

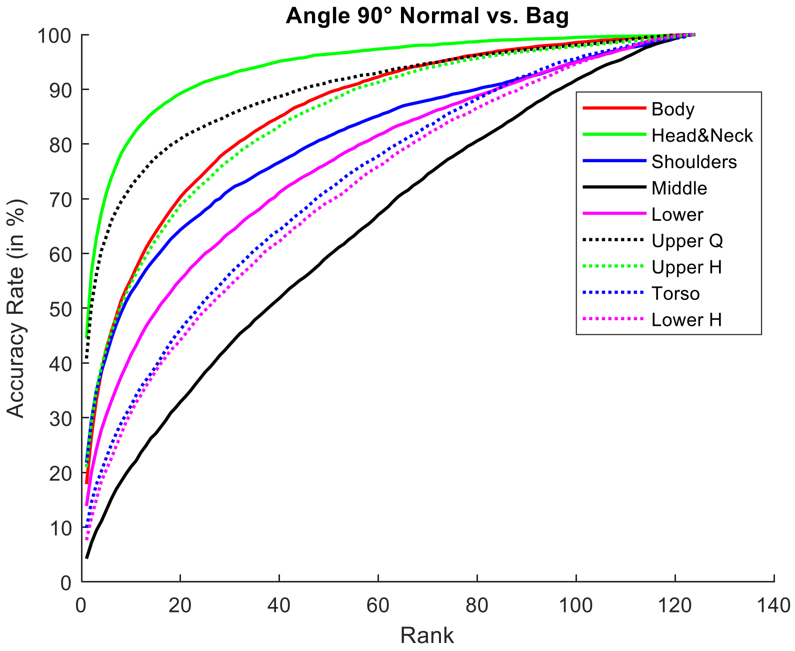


Figure 4.15: CMC curves of the cross-scenario classification showing the results of the Normal Scenario against the Bag Scenario from 90° view

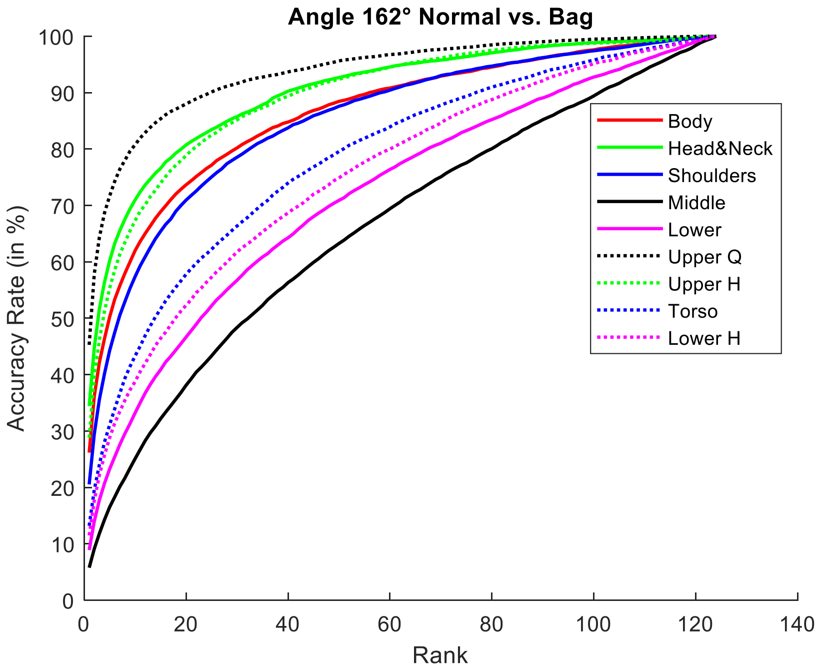


Figure 4.16: CMC curves of the cross-scenario classification showing the results of the Normal Scenario against the Bag Scenario from 162° view

Comparing with Normal vs. Normal scenario and considering the accuracy rates presented in Table 4.4, the performance was negatively affected by wearing the bag, most significantly, in the middle parts of the body or the segments that contain the *Middle* segment, such as the *Body*, *Upper H*, *Middle*, *Torso,* and *Lower H*. The segments that were least affected by carrying the bag were *Head & Neck*, *Shoulders*, *Upper Q,* and *Lower* as shown in Table 4.4. The *Upper Q* segment outperformed other segments in the straight views, but the *Head & Neck* segment outperformed in the side views. In the Normal versus Normal scenario, the *Lower* segment was the least accurate segment. The *Middle* segment resulted in the lowest accuracy rates in the Normal versus Bag scenario.

To summarise, carrying a bag directly impacts body shape, especially the middle parts of the body, thus the results significantly show lower levels of accuracy in all body parts in contrast with Normal scenario. Comparing the *Lower* and *Lower Half* segments results in Normal vs. Normal scenario with Normal vs. Bag scenario shows a slight decrease of accuracy. This means that carrying a bag indirectly influences the dynamical feature of the shape of these segments.

Table 4.4: First rank of accuracy rates (in percentage) for Normal vs. Bag classification showing the results of all segments at all views

Segment Angle	Body	Head & Neck	Shoulders	Middle	Lower	Upper Q	Upper H	Torso	Lower H
0°	39	27	22	11	11	**48**	40	21	19
18°	23	25	18	07	07	**34**	27	15	11
36°	18	24	13	04	07	**32**	23	11	2
54°	15	**29**	11	04	09	**30**	19	8	7
72°	18	**40**	17	04	11	35	21	9	7
90°	18	**44**	22	04	14	41	21	1	8
108°	16	**42**	17	04	10	39	18	09	5
126°	15	**38**	15	04	08	**37**	2	09	6
144°	19	39	01	05	08	**46**	27	11	7
162°	26	34	20	06	09	**45**	29	13	11
180°	39	35	18	10	15	**58**	39	20	2

4.5.2 Normal vs. Clothes

The classifier is trained on Normal scenario sequence and tested on Clothes scenario sequence, both are from the same viewing angle. Figure 4.17- Figure 4.20 show the CMC curves of all segments from four views, including 0°, 54°, 90°, and 162° respectively. Also, Table 4.5 shows the first rank of the accuracy rate of each segment in all views.

In general, the performance of the Normal versus Clothes scenario is negatively affected by the scenario change more than the Normal versus Bag scenario. The general results depicted in Table 4.5 do not show any pattern in terms of the most or least accurate segments. This indicates that extracting the static variation of the shape description employing the GFD descriptor is not a comparable way to overcome changes in the scenario with the subject wearing heavier clothes.

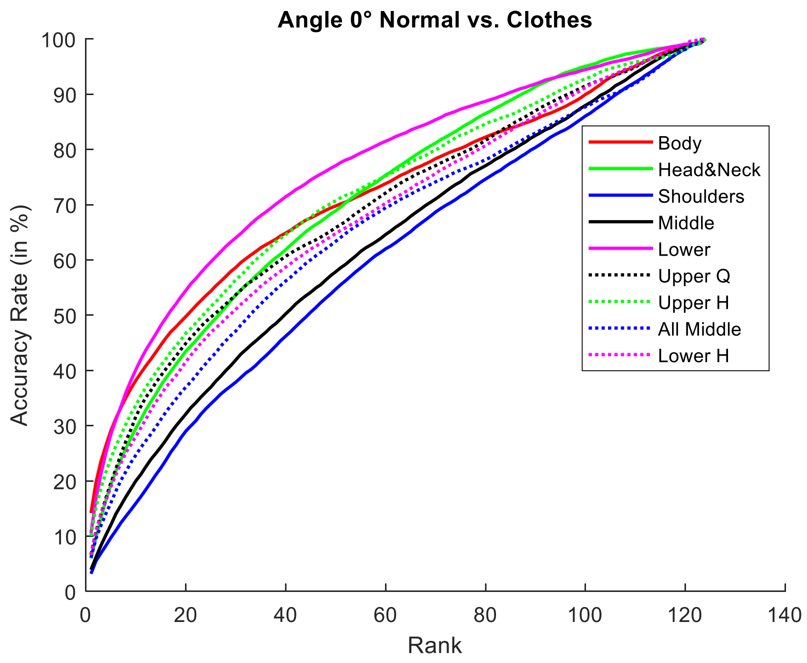


Figure 4.17: CMC curves of the cross-scenario classification showing the results of the Normal Scenario against Clothes Scenario from 0° view

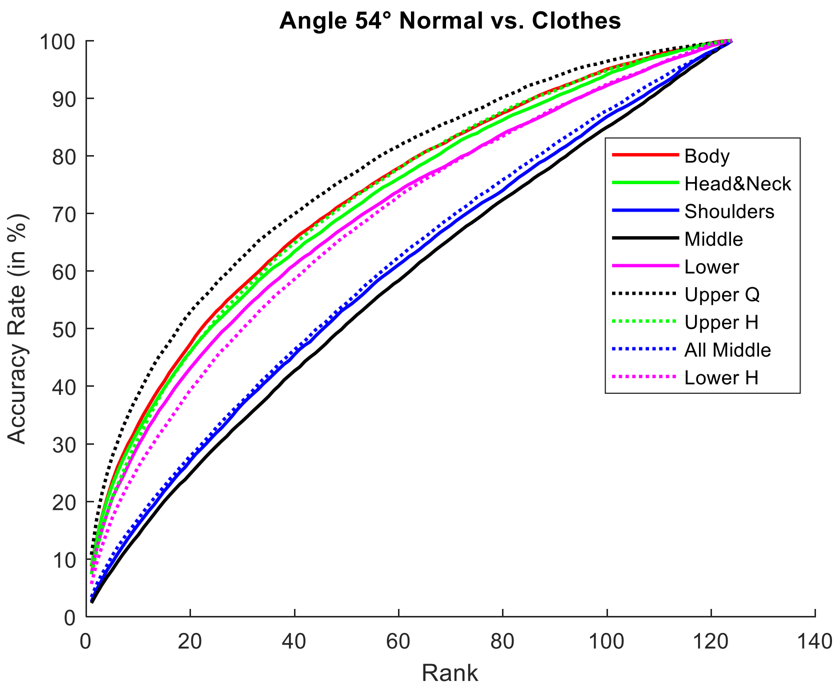


Figure 4.18: CMC curves of the cross-scenario classification showing the results of the Normal Scenario against Clothes Scenario from 54° view

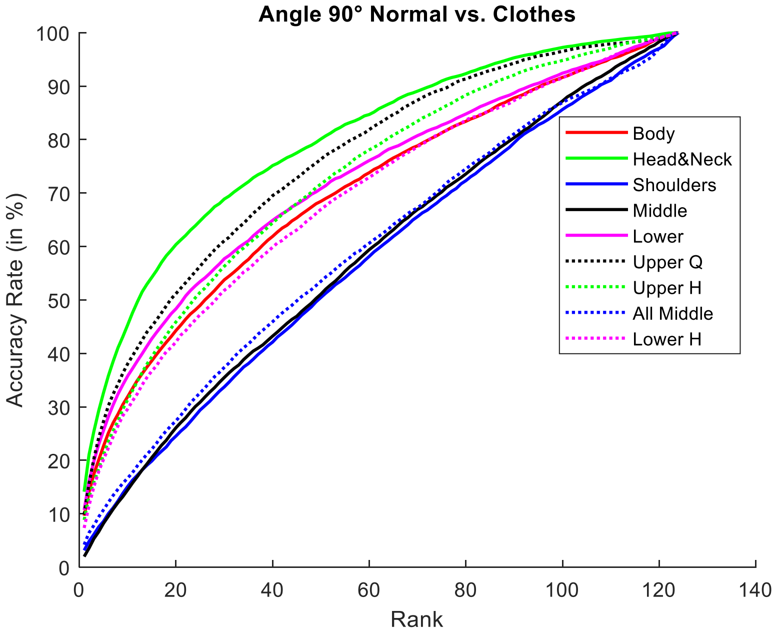


Figure 4.19: CMC curves of the cross-scenario classification showing the results of the Normal Scenario against Clothes Scenario from 90° view

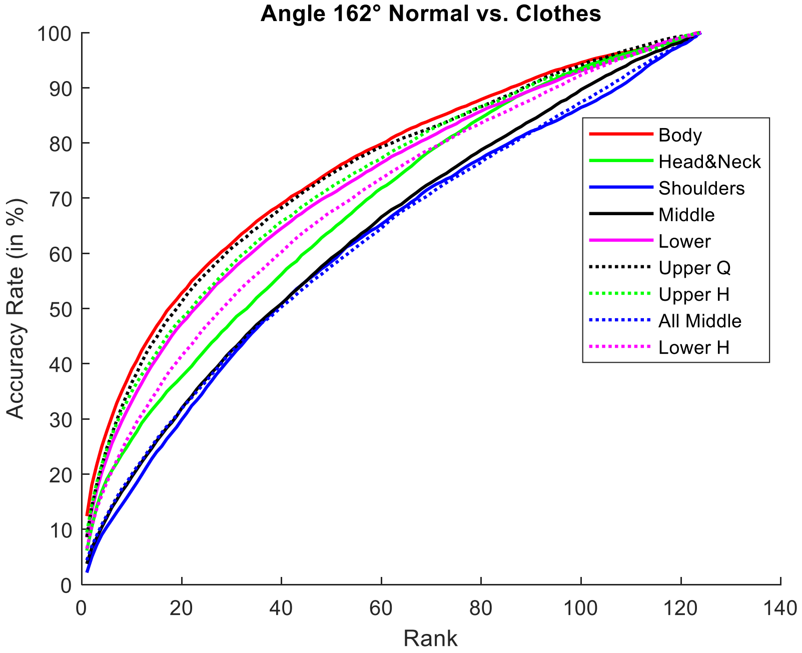


Figure 4.20: CMC curves of the cross-scenario classification showing the results of the Normal Scenario against Clothes Scenario from 162° view.

Table 4.5: First rank of accuracy rates (in percentage) for Normal vs. Clothes classification showing the results of all segments at all views

Segment **Angle**	**Body**	**Head &**	**Shoulders**	**Middle**	**Lower**	**Upper Q**	**Upper H**	**Torso**	**Lower H**
0°	**14**	6	3	4	**10**	7	1	5	6
18°	**10**	6	2	3	7	8	**9**	4	6
36°	**9**	7	3	3	7	**9**	7	4	6
54°	**9**	9	2	2	7	**10**	7	3	6
72°	9	**12**	2	2	10	**11**	8	3	7
90°	**10**	**14**	3	2		**10**	8	4	7
108°	8	**11**	3	2	8	**11**	8	3	6
126°	**9**	8	3	2	7	**12**	**9**	3	6
144°	**1**	8	3	3	7	**12**	**1**	4	7
162°	**12**	61	2	4	8	**9**	**9**	4	6
180°	**17**	6	3	6	**14**	6	11	6	11

4.6 Summary

This chapter demonstrates that body shape description can be utilised for person re-id. The results reflect that shape description delivers higher levels of discrimination when the subject is recorded from straight views than from side views. The main factors that impact the results are body orientation and shape variation caused by hand and leg shape.

Investigating the performance of different segments of the same body showed variation in segment performance. In addition to the entire body shape descriptor, the *Head & Neck* segment proved a comparable segment, outperforming all segments in the side views.

These results encouraged the integration of segments, where two or three connected segments are joined together. The integration performance indicated improvement, even in the segments that contained high shape variations, such as the *Lower* and *Middle* segments, caused by the subject's motion. Apart from the *Lower* segment, all segments performed better in the straight views than the side views, mostly due to body orientation within the same sequence.

Examining the inter-view situation showed that accuracy was negatively affected. However, the results also reflected that the more the tested view was close to the training view, the more the performance was improved, and vice versa. For example, if the training view is 90° then testing the viewing angles 72° and 108° showed higher performance than the other views. Because the body orientation in 72° and 108° is the most similar to the body orientation in 90°

The scenarios that impact body shape while a subject is in transition were investigated, specifically, carrying a bag and wearing heavy clothes. The results indicated that the bag had less impact on the shape description uniqueness than wearing heavy clothes. There is a need to examine other significant scenarios that affect the body shape, such as wearing hats and sunglasses.

Generally speaking, extracting and utilising shape static variation through shape description reflects comparable performance, yet a number of weaknesses occurred. In order to overcome these weaknesses in recognising the same subject in a situation when they wear heavier clothes, we will need to find a different way to use shape description more than the static variation of the description.

Chapter 5

Body Dynamic Variation Using Dynamic Time Wrapping

5.1 Introduction

The experiments conducted in Chapter 4 illustrated the effectiveness of using the body and body segments' shape descriptor using a Generic Fourier Descriptor (GFD) in person re-identification (re-id). The classification process was implemented frame-by-frame. These experiments investigated the static variation of the shape-based person re-id, which refers to the changes that appear on the shape descriptor on a frame-by-frame basis. Thus, this approach represents image-based person re-identification.

In the description provided in this chapter, similar factors are examined for a similar purpose; however, the experiments discussed here are designed to examine the dynamic rather than the static features of the body shape. The dynamic features focus on body shape descriptors while the subject is in motion. That is, the dynamic variation represents the changes that occur on the body shape based on a whole sequence, where multiple frames are required. Thus, this approach represents video-based person re-id.

The results of the analysis of image-based person re-id focusing on static features of the body GFD shape description were illustrated in Chapter 4 utilising the Linear Discriminant Analysis (LDA) classifier. The results indicated the potentially superior performance of the proposed approach. However, some aspects of that process still require improvements, such as accommodating changes in the scenario or in the viewing angle.

In this Chapter, results are presented from a number of experiments conducted to assess the performance of dynamic features of the body and body segments GFD shape description for person re-id. The classification algorithm used in the experiments described in this chapter is Dynamic Time Wrapping (DTW). This is followed by a detailed analysis of the data to identify the factors that directly and indirectly influence segment performance.

This chapter is organised as follows: Section 5.2 describes the methodology used to implement the shape-based person re-id matching process based on the shape description dynamic feature using the DTW algorithm. Section 5.3 presents the initial results, including the accuracy rates, of performing the DTW on the entire body shape as well as on the proposed body segments. Also presented in this section is an analysis of the results designed to identify the factors that affect the accuracy rates of each segment, including viewing angle and body orientation. Section 5.4 presents the performance data of the inter-view classification approach, where different viewing angles are assigned to the training and test sets. Section 5.5 illustrates the accuracy rates of the matching process from two different scenarios, namely, Normal versus Bag and Normal versus Clothes, see 3.3.1. Section 5.6 compares the performance of the proposed method with that of other related state-of-art person re-id techniques. Finally, Section 0 summarises the implementation aims, methodologies used, and results obtained as presented in this chapter.

5.2 Methodology

This section describes the second methodology proposed for matching subjects based on their shape description using the shape description dynamic feature. In the shape description dynamic feature, temporal variations on the body shape description are considered in the classification step, performed on the entire sequence at once (i.e., one sequence is composed of multiple frames). Therefore, the dynamic feature approach examined in this method represents video-based person re-identification. The classification procedure is executed using the DTW algorithm, which calculates the similarity between two temporal sequences. It returns a distance scalar of the two sequences' optimal alignment [105]. Figure 5.1 shows the proposed system framework of this research, highlighting the focus of the analysis presented in this chapter.

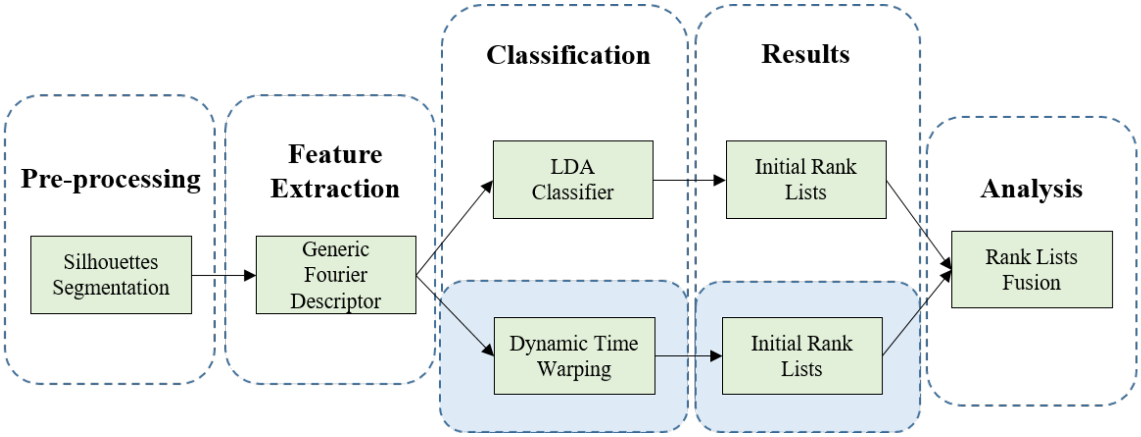


Figure 5.1: Proposed system framework highlighting the focus on body shape dynamic variation using DTW

5.2.1 Dynamic Time Wrapping Implementation

In the previous experiment presented in Chapter 4, the GFD shape descriptor was extracted from each frame, and the classification step was implemented using the LDA classifier, which is based on a frame-by-frame or image-based approach. Consequently, a rank list was generated for each frame. However, in order to detect the unique shape dynamic feature between subjects using the DTW algorithm, the whole sequence of frames is required (i.e., one sequence is composed of multiple frames). Therefore, the matching step of the process using the DTW algorithm will be implemented as a sequence-by-sequence or video-based scheme. Therefore, one rank list will be generated for the entire sequence. The underlying hypothesis of using DTW in this experiment is that the similarity of the dynamic variation between genuine identities should be greater than the similarity of the dynamic variation between imposter identities.

5.2.1.1 Single-Dimension Dynamic Time Wrapping

The original purpose of DTW was to find the similarity of dynamic features between two single-dimension (1D) feature vectors representing two sequences. The 1D version of the DTW can be calculated as follows: suppose that the input feature vectors are vector $A_a(i)$, where $i = 1, \dots, n$, and vector $B_b(j)$, where $j = 1, \dots, m$, and n and m represent the number of elements in the vectors A and B, respectively. The DTW algorithm, then, calculates the distance as the minimum distance from the beginning of the DTW table to the current position (i, j). The DTW table can be defined as follows:

$$DTW(A, B) = D(i, j) = d(i, j) + min\begin{cases} D(i-1\,, j) \\ D(i, j-1) \\ D(i-1, j-1) \end{cases} \qquad (\,5.1)$$

where $D(i, j)$ is the cost node associated with $A(i)$ and $B(j)$ as defined in

$$d(i, j) = (A[i] - B[j])^2 \qquad (\,5.2)$$

Table 5.1 shows the detailed single-dimension DTW algorithm for finding the scalar distance between two sequences A and B.

5.2.1.2 Multi-dimensional Dynamic Time Wrapping

The aim of this experiment was to match identities of individuals based on the similarity of the dynamic variation in their shape description. The DTW algorithm was used to implement this process. As noted, the DTW algorithm was originally designed for 1D sequences. However, a multiple number of frames were present in the sequences for each subject in the examined dataset, and the number of frames in a sequence was different from one subject to the next. Each frame had an individual GFD feature vector length of 156 elements, as justified in 4.2.2. Accordingly, each sequence in the examined dataset was considered a multidimensional sequence, as each consisted of multiple feature vectors (i.e., one feature vector for each frame).

Table 5.1: Single-Dimension DTW Algorithm

Single-Dimension DTW Algorithm

```
Input: $A = [a_1, \ldots, a_n]$
Input: $B = [b_1, \ldots, b_m]$
Let $d$ be a distance between coordinates of $A$ and $B$
Let $D$ be a cost matrix
for $i = 1$ to $n$ do
    for $j = 1$ to $m$ do
        $d(i,j) = (A(i) - B(j))$^2
    end for
end for
$D(1,1) = d(1,1)$
for $i = 2$ to $n$ do
    $D(i,1) = d(i,1) + D(i-1,1)$
end for
for $j = 2$ to $m$ do
    $D(1,j) = d(1,j) + D(1,j-1)$
end for
for $i = 2$ to $n$ do
    for $j = 2$ to $m$ do
        $D(i,j) = d(i,j) + \min\begin{cases} D(i-1,j) \\ D(i,j-1) \\ D(i-1,j-1) \end{cases}$
    end for
end for
Return $\sqrt{D(n,m)}$
```

Therefore, the one-dimensional DTW needed to be modified to handle multidimensional sequences. There are a number of ways to implement multidimensional DTW [137]. However, the main goal here was to examine the dynamic feature of the shape description on the video-based approach. To achieve that goal, the implementation of such a concept can be defined as follows: Suppose there are two sequences, S_1 and S_2, consisting of x and y number of shape descriptor feature vectors, respectively, each of which represents one frame of the sequence. In this scenario, S_1 and S_2 are defined as:

$$S_1 = \begin{bmatrix} A_1 \\ A_2 \\ . \\ . \\ A_x \end{bmatrix} \text{ and } S_2 = \begin{bmatrix} B_1 \\ B_2 \\ . \\ . \\ B_y \end{bmatrix} \qquad (5.3)$$

where $A_x = [a_1, \dots, a_k]$ and $B_y = [b_1, \dots, b_k]$ and k is the length of the extracted GFD feature vector. The equation necessary to wrap these two multidimensional sequences using multidimensional DTW can be defined as follows:

$$\sum_{i=1}^{k} DTW\left(S_1(i), S_2(i)\right) \qquad (5.4)$$

Multidimensional DTW can be described as the accumulative distance of the DTW of the k^{th} dimension of all feature vectors from S_1 and S_2. The theoretical explanation of this implementation can be described as follows: two sequences can be wrapped by generating two vectors, one from each sequence. One vector is generated by horizontally concatenating the k element of all GFD feature vectors of one sequence. The generated vectors are then fed to the DTW algorithm. This process goes through all the elements of the two sequences' vectors. The resulting distances are accumulated to arrive at one scalar distance for the two sequences. This means that the DTW alignment of two sequences consumes k iterations.

There are two reasons to implement this approach of multidimensional DTW, the first of which is computation time. The proposed concept consumes k iterations, where k in this research is 156, as justified in 4.2.2. The other multidimensional DTW concept (not implemented in this research) is implemented by comparing one frame GFD feature vector with another frame GFD feature vector. There are two issues to consider when implementing this concept. First, this concept works on the image-based aspect, wherein this part of the research discovers the shape description using the video-based approach. Second, the consumption time for this concept can be defended as: n * m, where n and m are the number of the frames in sequence one and two, respectively. The average number of the frames in a sequence in the examined dataset (i.e., CASIA Dataset B) is 68 frames. Consequently, the average consumption time of wrapping two sequences applying this concept is 68*68 = 4,624 iterations per two sequences.

The other fundamental reason behind implementing the proposed multidimensional DTW concept is that the length of the GFD feature vector for each frame is fixed. Consequently, the dynamic variations between one sequence and the other are lying in each vertical dimension of the sequence (i.e., one dimension of the sequence is the k^{th} element of all the feature vectors of that sequence).

5.2.1.3 DTW Performance Evaluation

The scalar distance of two multidimensional sequences is found by implementing multidimensional DTW as clarified in the previous section. Subsequently, a rank list must be generated for each sequence. This is accomplished in two steps. Step one is to find the distance between all the sequences in the dataset and store them in the *Sequences Distances Matrix*. Note that each subject has two different sequences that were recorded at different times, which was assigned as training and test sets.

Once the *Sequences Distances Matrix* is calculated, the second step is to generate the rank list of each sequence based on the distance provided between this sequence and all other sequences. As the distances are ascendingly ordered, then these distances are replaced with their corresponding sequences' identities, which generates the identities rank list.

The rank lists are then used to produce the Cumulative Match Curve (CMC) curves. The CMC curve is the evaluation metric commonly used in person re-id research; it was also used in the first experiment of this research. The CMC curves are produced by calculating the number of correct matches in each rank of all rank lists. Formally, CMC curves represent the accuracy rates of the implementation of DTW on the GFD feature vectors to find the dynamic feature similarity between subjects' sequences. The CMC curves are presented in the next section.

5.3 Body and Body Segments Dynamic Feature

As noted, each subject in the CASIA Dataset B was provided with two different sequences, recorded under the same conditions but at different times. From one subject to another, the sequences varied in terms of the number of frames. In the process presented in Chapter 4, the GFD shape descriptor was extracted from each frame, forming an individual feature vector for each. In this part of the study, all feature vectors of both sequences of all subjects in the dataset were sorted through the DTW algorithm. The main goal here was to assign an identity number to each sequence of the test set. This number assigned was determined by calculating the similarity, using the distance between each sequence in the test set and training set, where they simulated the watch list and the enrolled subjects, respectively, in a real CCTV intelligent person re-id system.

The DTW algorithm determined the distance between two multidimensional sequences based on the temporal and dynamic feature in their GFD feature vectors. This aspect of the process is important to this research, as the aim is to investigate the role of utilising shape description as an identifier in person re-id applications, since the body shape is prone to significant changes when the subject in motion.

5.3.1 Body Performance

The sequences in the dataset were aligned by DTW based on their dynamic variation where the aligned sequences are of the same viewing angle. As above-mentioned, the dataset provided two sequences for each subject to be recorded under the same conditions. One of them was used as in training set and the other one was used as in the test set. This was implemented on the normal scenario sequences; the results of the entire body segment are presented in Figure 5.1. The value of the first rank from each viewing angle can be found inthe side views.

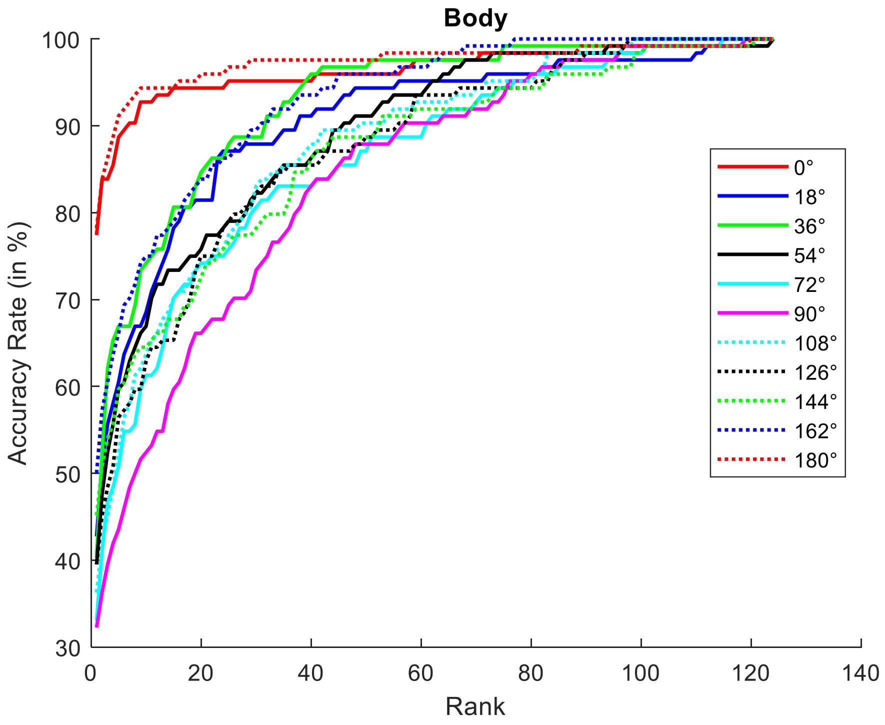


Figure 5.2: CMC curves of the entire body shape accuracy rates based on DTW from 11 viewing angles

The CMC curves presented in Figure 5.2 show the performance of matching the identities based on their dynamic variation: each curve presents one viewing angle. The most obvious finding is that there was a unique dynamic feature between the subjects' shape descriptions. In addition, as the static variation of the shape description feature vectors showed variability from one viewing angle to another (as explained in Chapter 4), the dynamic feature in the shape description also fluctuated between viewing angles of the same body. This means that the accuracy rates of DTW of the body shape description from 0° are different from those at 90° and so on.

As the dynamic variation of the body shape description showed comparable accuracy rates, further experiments were conducted to explore the dynamic feature performance of the proposed segments. The next section discusses this aspect of shape-based person re-id research.

5.3.2 Segments Performance

Experiments like the one conducted to examine body dynamic variation were performed on the proposed body segments. In Chapter 3, the body silhouette was segmented into four non-overlapped segments. Then, further segments were proposed in Chapter 4. These segments were discussed in detail in Sections 3.4.2 and 4.3.3. The proposed segments are *Head & Neck*, *Shoulders*, *Middle, Lower, Upper Quarter* (*Upper Q*), *Upper Half* (*Upper H*), *Torso,* and *Lower Half* (*Lower H*). Table 5.2 provides the accuracy rate of the first rank of the dynamic variation of all proposed segments from all viewing angles using DTW, where the aligned sequences are of the same viewing angle. In this Table, the highest accuracy rates are bold. Figure 5.3, Figure 5.4, Figure 5.5 and Figure 5.6 show the accuracy rates of these body segments from four different viewing angles 0°, 54°, 90° and 162° respectively. In the next subsection, the main findings for each segment are presented.

Table 5.2: The accuracy rate of the first rank (in percentage) of the dynamic variation of all proposed segments from all viewing angles using DTW

Segment **Angle**	**Body**	**Head & Neck**	**Shoulders**	**Middle**	**Lower**	**Upper Q**	**Upper H**	**Torso**	**Lower H**
0°	77	42	50	50	31	73	**86**	62	61
18°	43	47	56	46	21	**65**	**63**	50	36
36°	40	45	47	47	30	**60**	**62**	47	43
54°	40	60	46	53	39	**63**	57	46	46
72°	33	62	56	45	30	**68**	60	54	34
90°	32	64	50	43	23	**68**	59	50	31
108°	36	57	52	44	27	**67**	62	56	36
126°	39	41	35	39	27	**52**	**52**	39	39
144°	45	45	40	43	31	**57**	**55**	39	44
162°	50	44	52	57	29	**69**	**70**	58	46
180°	78	47	50	60	39	75	**84**	70	67

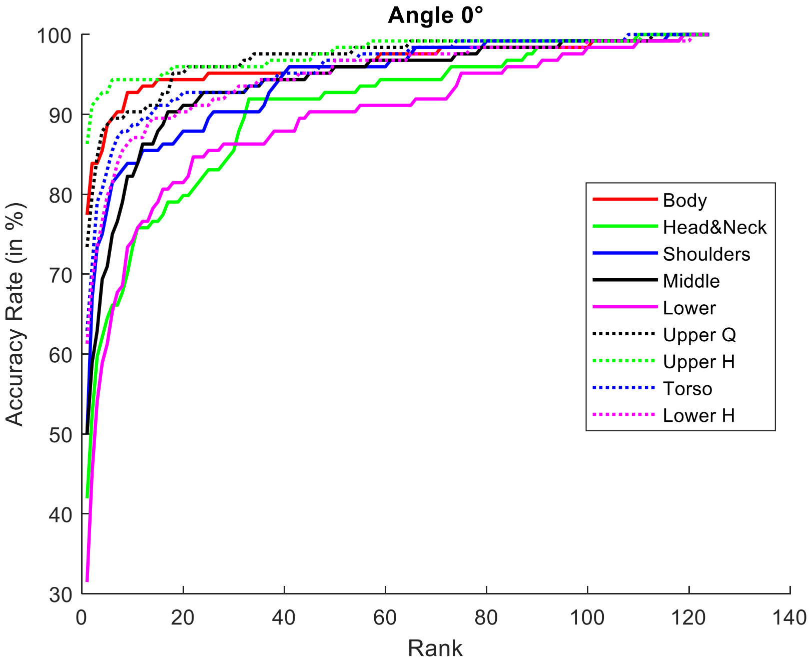


Figure 5.3: CMC curves of the accuracy rates of different body segments, where the aligned sequences of the 0° angle

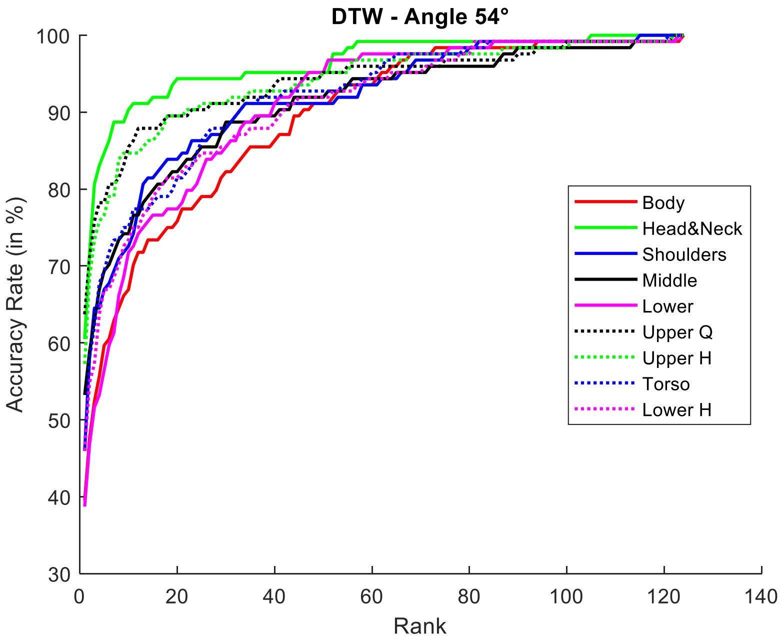


Figure 5.4: CMC curves of the accuracy rates of different body segments, where the aligned sequences of the 54° angle

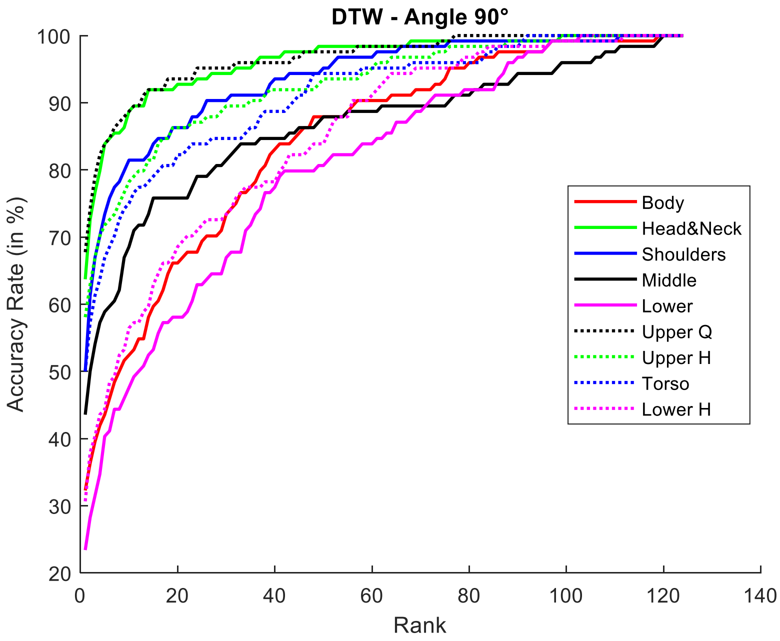


Figure 5.5: CMC curves of the accuracy rates of different body segments, where the aligned sequences of the 90° angle

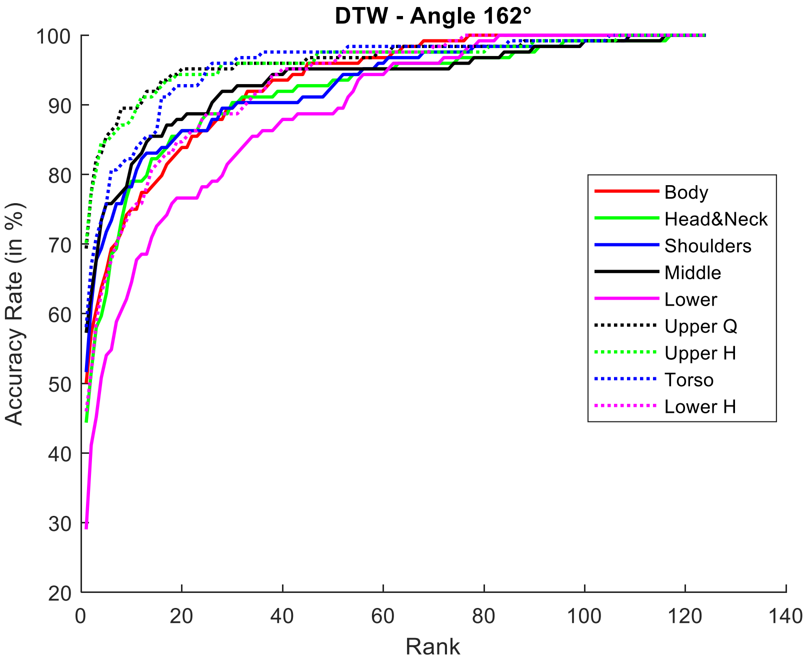


Figure 5.6: CMC curves of the accuracy rates of different body segments, where the aligned sequences of the 162° angle

Head & Neck; the CMC curves that show the ranks of all viewing angles demonstrate that this segment was one of the outperforming segments in the side view (i.e., 90°) and in the views that primarily present the side part of the segment, such as 54°, 72°, and 108° angles. This performance is compared to the straight views (i.e., frontal or back views, including 0° and 180°) or the views that tend to capture straight views more than side views, such as 18°, 36°, 144°, and 162°. Comparing this segment performance with other segments, *Head & Neck* was close to the outperforming segment (i.e., *Upper Q*) with 3%, 5%, and 4% differences from the views 54°, 72°, and 90°, respectively.

Shoulders; no pattern emerged from the results for this segment. The accuracy rate of the first rank fluctuated from one view to another. The straight and side views—0°, 90°, and 180°—were very similar at 50%, 50%, and 52%, respectively.

Middle; similar to the *Shoulders* segment, the *Middle* segment results revealed no pattern. Although the *Middle* segment size is larger than the *Shoulders* and *Head & Neck* segments, it is considered one of the body parts that changes shape most often while the subject is in motion. Therefore, DTW was not able to find the uniqueness of the shape dynamic feature between the subjects' shape description features. The only segment that the *Middle* segment outperformed in all situations was the *Lower* segment.

Lower; compared to the rest of the segments, this segment maintained the lowest accuracy levels in all angles. The *Lower* segment is the other part of the body that changes shape most often while the subject is in motion. Therefore, DTW was not able to determine the uniqueness of the shape dynamic variation in between the subjects' shape description features.

Upper Quarter; *Upper Q* segments gained the highest first rank accuracy rates in most of the viewing angles, except for the straight viewing angles (i.e., 0° and 180°), where the higher accuracy was achieved by the *Upper H* segment. One reason for this performance may be that this segment is composed of the parts of the body (head, neck, and shoulders) that are most stable while the subject is in motion. Consequently, the influence of motion on the shape descriptions was considerably less than it was on the other segments that contain parts that are less stable during motion. This aids the DTW in finding the unique dynamical shape variation between subjects.

Upper Half; the *Upper H* segment showed high accuracy levels compared to the other segments as well. It outperformed all segments in the straight viewing angles (i.e., 0° and 180°) with 86% and 84% first rank accuracy rates, respectively. In addition, the first rank accuracy rates were close to the outperforming segment *Upper Q* from the viewing angles that slightly captured the side of the body, such as 18°, 36°, 144°, and 162°, with an average difference of 2%.

Torso; this segment covers the whole arm and hand shape of the body. The accuracy rates of the *Torso* segment in the straight viewing angles (i.e., 0° and 180°) outperformed the other views.

Lower Half; this segment is one of the lowest performing segments, followed by the least accurate segment *Lower*. This mostly due to the inclusion of most moving part which is the leg. Yet, the accuracy levels of the straight views of the *Lower Half* segment tend to be higher than the side views.

5.3.3 Further Analysis of Proposed Angle-based Approach with DTW

In the previous section, the general performance of each segment was presented, whereas in this section, the performances of all segments are compared in terms of the main influencer factor/s. all the presented results are implemented as an angle-based approach, where the aligned sequences are of the same viewing angle.

First, there is a common finding on the first rank accuracy rates of the segments identified by the DTW algorithm. The arm, hand, and leg parts of the body were the parts with a large impact on the body shape while the subject was in motion. At the same time, the motion influence of these parts had more impact on the silhouettes captured from the side views (i.e., 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, and 162°) than from the straight views (i.e., 0° and 180°). Therefore, the stability patterns between the segments comprising arm, hand, and leg parts were similar. The visual results of the first rank accuracies of the segments *Body*, *Upper H*, *Torso,* and *Lower H* in Figure 5.7 and

Figure 5.8 show that the first and last views performed better than the other views. The first and the last views represent the straight views. However, the *Upper H* segment, in fact, outperformed these segments in all the viewing angles. One justification might be that the *Upper H* segment eliminates the most movable part of the body—the *Lower* segment. In addition, it contains the most stable parts of the body, which are the head, neck, and shoulders.

An analysis of the accuracy rates of the *Body, Upper H, Torso,* and *Lower H* segments displayed in Figure 5.7 and Figure 5.8 from a different perspective reveals performance dramatically decreased when comparing the straight views with the rest of the views. This confirms that the segments between the straight views were highly influenced by the viewing angle factor.

Results of the *Upper Q* segment accuracy rates in Figure 5.7 (i.e., *Head & Neck* and *Shoulders*) and Figure 5.8 confirm that this segment was influenced by a different factor. The accuracy rates are mostly in the similar level in all angles, except 36°, 54°, 126° and 144° angles. These viewing angles, in particular, are the most angles that contain high level of body orientation, where the body direction captured in one sequence get change over the time.

As noted previously, the *Upper H* was the outperforming segment in all views followed by the *Upper Q* segments in the the side views. However, the accuracy rates of the *Head & Neck* segment were highly comparable to the *Upper Q* segment, especially in the side views (i.e., 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, and 162°). This segment, in fact, is less likely to be occluded by another subject compared to other segments of the body. As a result, the likelihood of correctly extracting this segment of the body is higher than for any other segment.

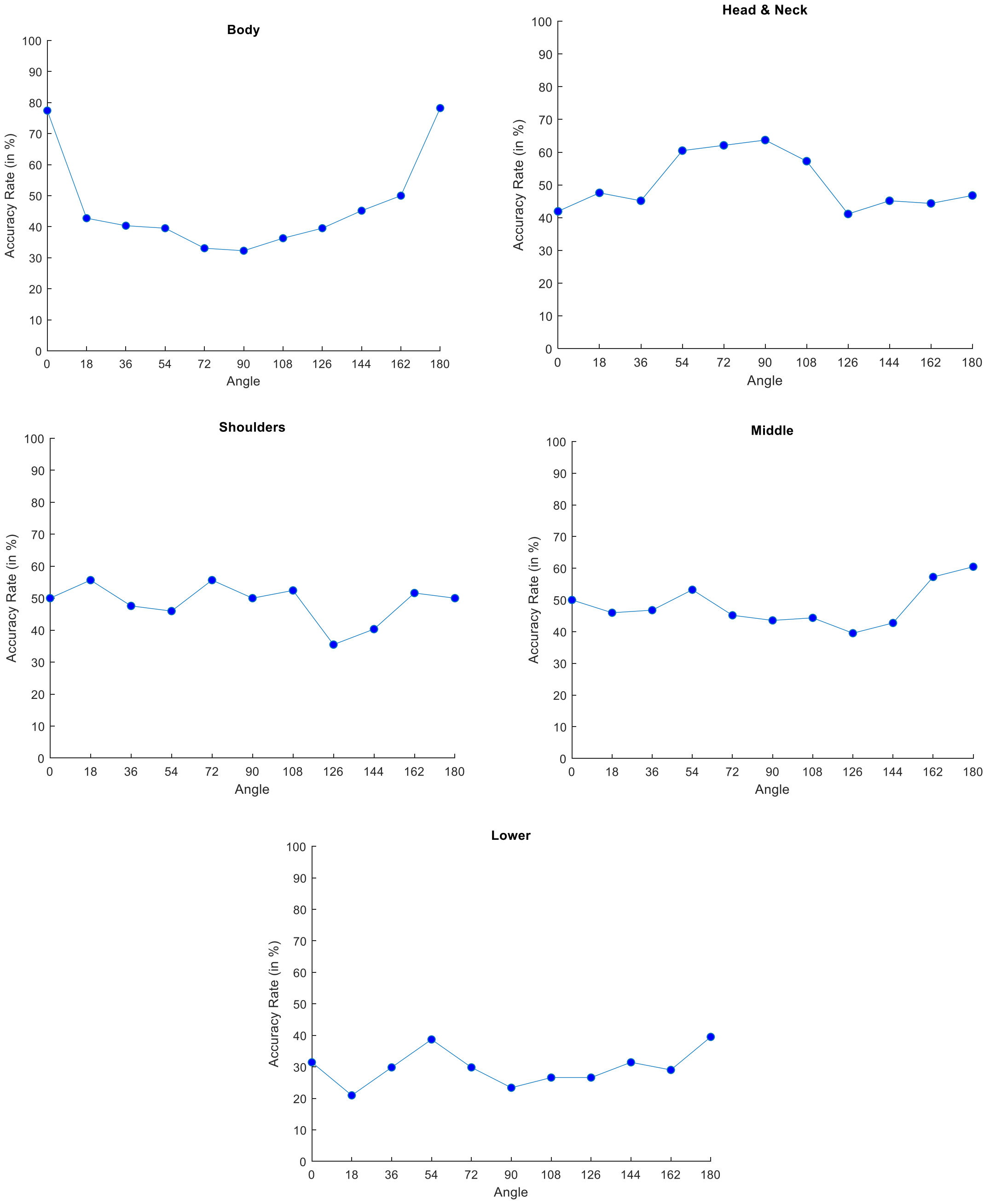


Figure 5.7: First rank accuracy rates from the 11 viewing angles of the original proposed body segments (Body, Head & Neck, Shoulders, Middle, and Lower)

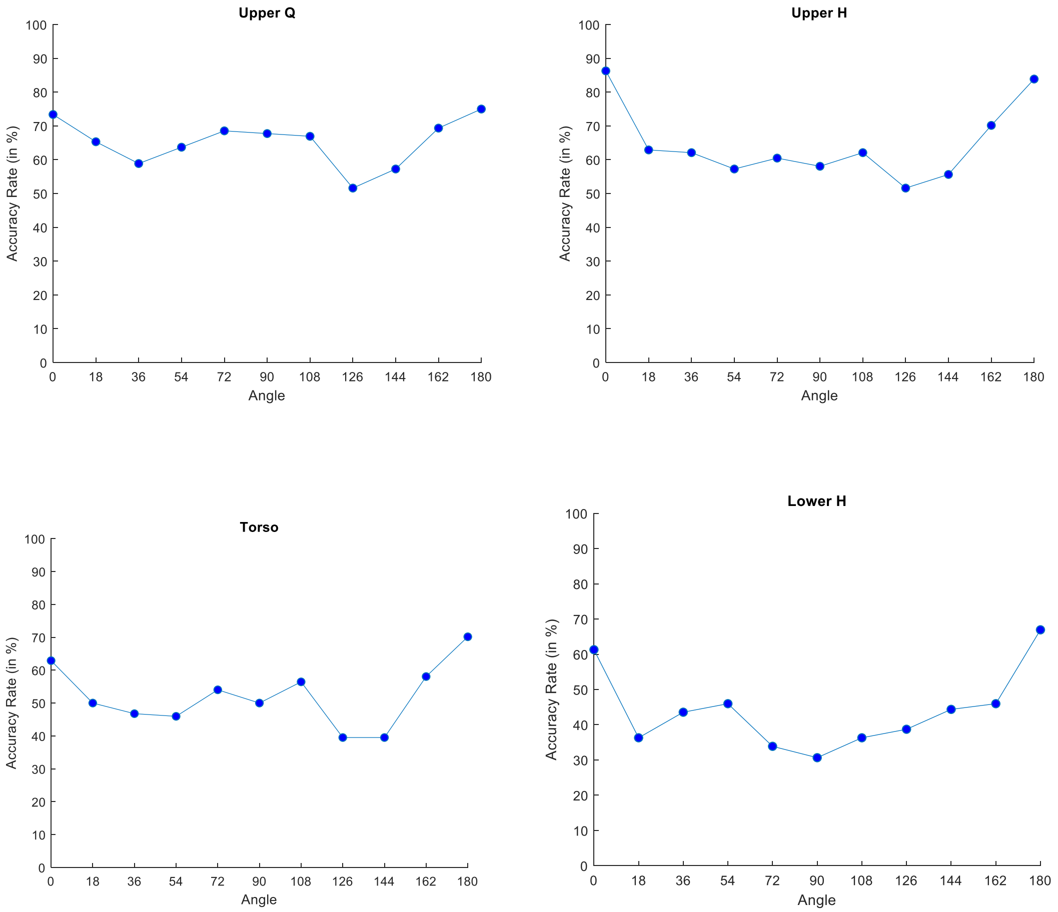


Figure 5.8: First rank accuracy rates from the 11 viewing angles of the additional proposed body segments Upper Q, Upper H, Torso, and Lower H segments

Comparing the performance of the *Torso* segment with the *Lower* segment involves comparing the accuracy of the arm and hand parts shape description with the leg shape description, which are non-overlapped segments representing the most movable parts of the body. Figure 5.7 and Figure 5.8 (*Torso* and *Lower* figures) and the numerical values of the first rank accuracy rates presented in Table 5.2 show that the shape description of the *Torso* segment (i.e., the arm and hand parts) was more discriminative than that of the *Lower* segment (i.e., leg parts), reflected in the average performances at first rank in the *Torso* segment (52%) and *Lower* segment (30%).

5.4 Inter-view Dynamic Performance

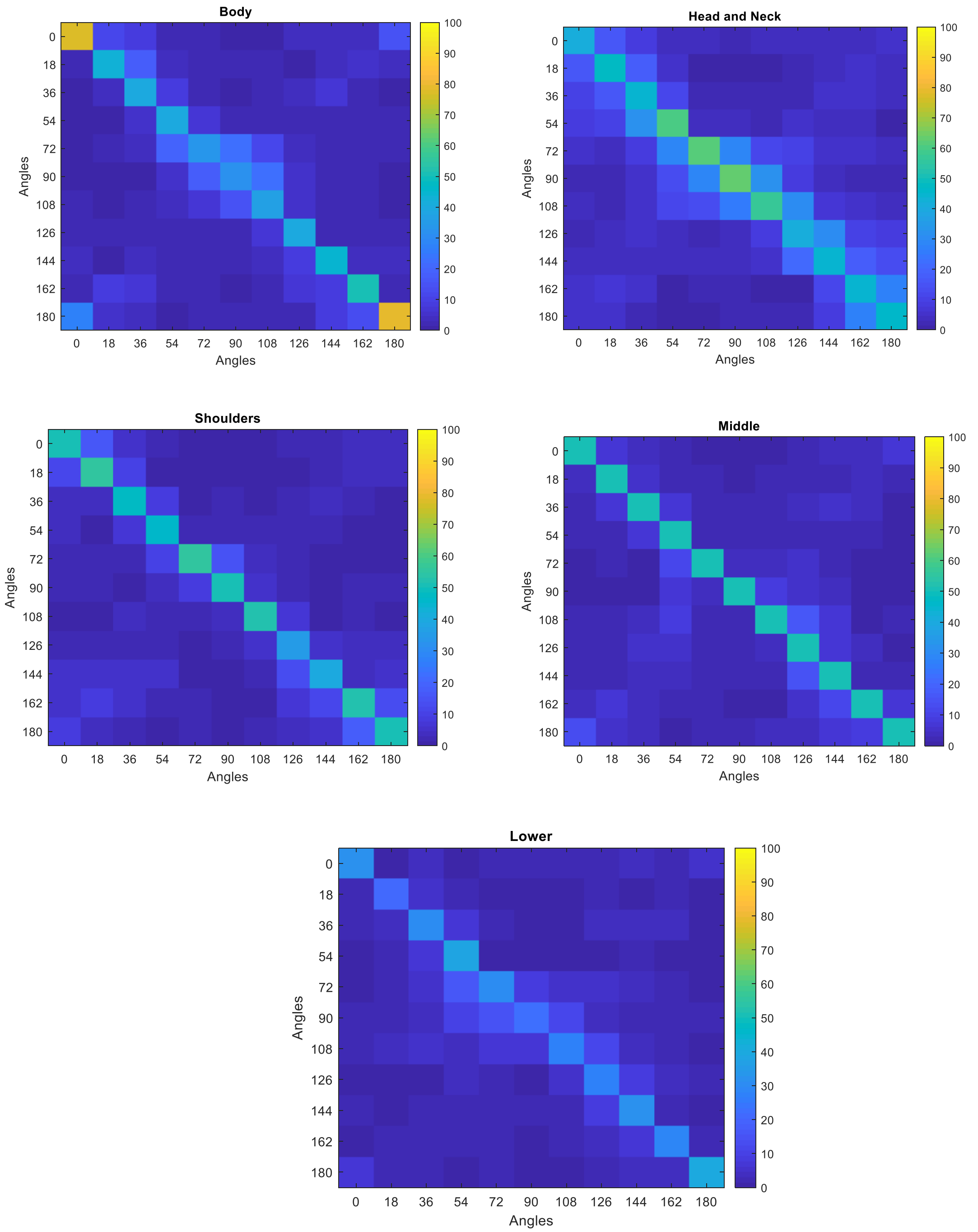
All experiments presented thus far in this chapter were designed to fall within the same viewing angle. Therefore, the DTW process involved wrapping two sequences from the same view. Chapter 4 described an experiment that was implemented based on the frame-by-frame approach (i.e., image-based), utilising shape description feature vectors, where the frames were from different views.

A similar experimental methodology was followed in this part of the research. The DTW algorithm was applied on the shape description feature vectors of two sequences from two different viewing angles. As above-mentioned, the dataset provided two sequences for each subject to be recorded under the same conditions. One of them was used as in training set and the other one was used as in the test set. The aim here was to investigate the performance of matching identities based on their shape description dynamic feature. This examined whether there was a unique dynamic variation between the subjects' shape descriptions from different viewing angles.

Figure 5.9 and Figure 5.10 visually show the first rank accuracy rates from implementing DTW on two sequences from different viewing angles. Each coloured map represents the result of a different segment. A diagonal in each figure indicates the best results compared with the rest of the figure, as it represents the results of two sequences from the same view.

Regardless of the same view matching performance shown in the diagonals, the Inter-view accuracy rates from using the DTW on the sequences' shape descriptors (i.e., video-based) are considered unacceptable when compared to results of the same viewing angles.

Comparing the results from the implementation of the DTW algorithm in the video-based approach with the results from the implementation of the LDA classifier in the image-based approach from Chapter 4 illustrates that the static variation built based on the frame-by-frame method performed slightly better than that of the video-based method implemented using the DTW. Accuracy within the same viewing angle was higher using the DTW on the video-based method than using the LDA with the image-based method.


Body

Head and Neck

Shoulders

Middle

Lower

Figure 5.9: Inter-views first rank accuracy rates from implementing DTW on two sequences from different viewing angles on the original proposed segments

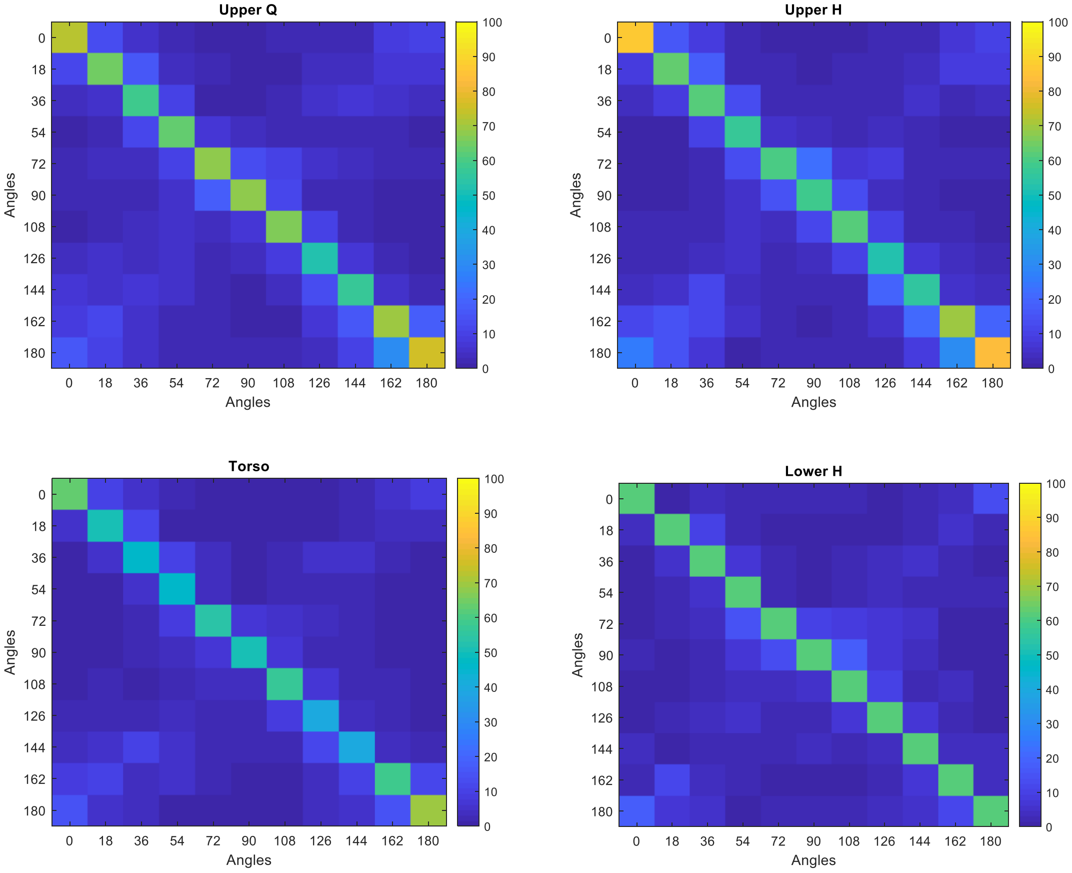


Figure 5.10: Inter-views first rank accuracy rates from implementing DTW on two sequences from different viewing angles on the additional proposed segments

5.5 Cross-Scenario Dynamic Performance

As discussed, real life scenarios in which people move about in public areas involve a significant amount of variation in multiple forms. One of these variations is appearance change, which can result from the presence or absence of heavy clothes, such as winter coats, and from the presence or absence of a bag, such as a handbag, backpack, or rucksack, carried by the subject. These changes, in fact, have direct and indirect impacts on the subject body shape. This leads to vital variations in the shape description. Therefore, the performance of DTW on two sequences of different scenarios required investigation.

The scenarios provided by the CASIA Dataset B are Normal, Clothes, and Bag scenarios. In the next section, an examination of the Normal versus Bag and Normal versus Clothes scenarios is presented.

5.5.1 Normal versus Bag Video Sequences

The DTW algorithm was implemented on all subjects sequences in the dataset; however, training sequence presents the subject in the Normal scenario, and the test sequence presents the same subject carrying a single shoulder bag, backpack or a handbag. The methodology followed in this experiment replicated implementation of the angle-based experiment as described previously in this chapter. Figure 5.11,

Figure 5.12, Figure 5.13 and Figure 5.14 show the cross-scenario CMC curves of all proposed segments from four angles: 0°, 54°, 90°, and 162° respectively. Table 5.3 illustrates the tabular values of the first rank accuracy rates of all proposed segments from all viewing angles. The highest accuracy rates in each angle is bold.

The results represented by the CMC curves and in Table 5.3 are considered, to some extent, low when compared to the results from the angle-based method with the Normal versus Normal approach. However, the accuracy of some segments, such as *Body*, *Upper H*, *Middle*, *Shoulders,* and *Torso,* dramatically decreased because the bag added to the shape of these segments, causing a vital change to the shape description.

The other segments, including *Head & Neck*, *Lower*, *Upper Q,* and *Lower H,* maintained a similar range of accuracy levels with only slight decreases. Although these segments were not directly affected by carrying a bag, the body pose in general was influenced by carrying the bag. This caused a noticeable variation on these segments' shape and, consequently, a moderately different shape description than the shape description of the segment in the Normal scenario.

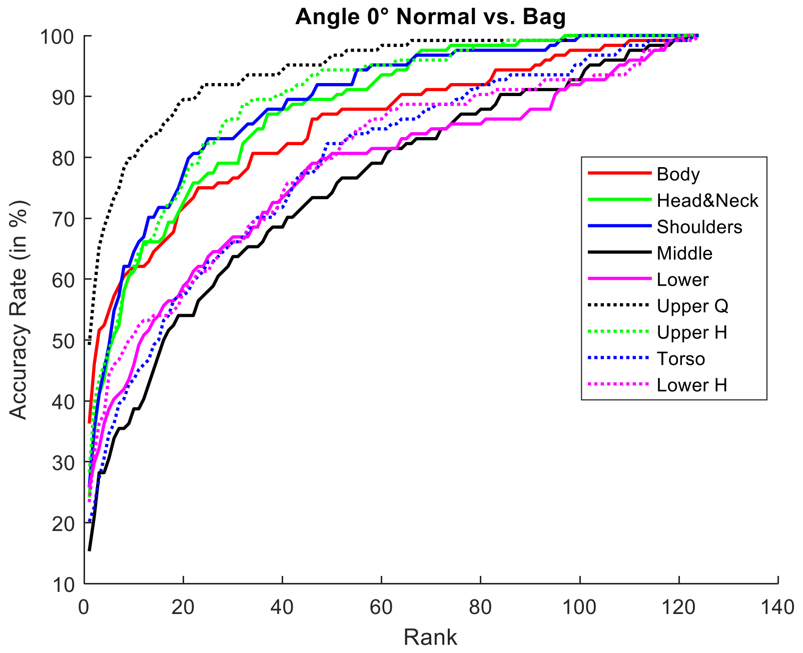


Figure 5.11: Normal versus Bag cross-scenario approach CMC curves based on implementation of DTW from Angle 0°

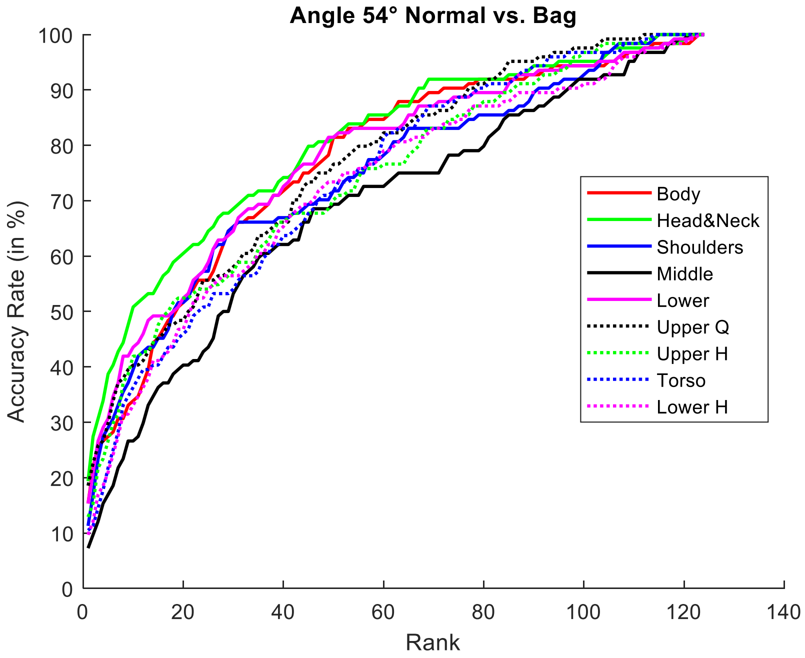


Figure 5.12: Normal versus Bag cross-scenario approach CMC curves based on implementation of DTW from Angle 54°

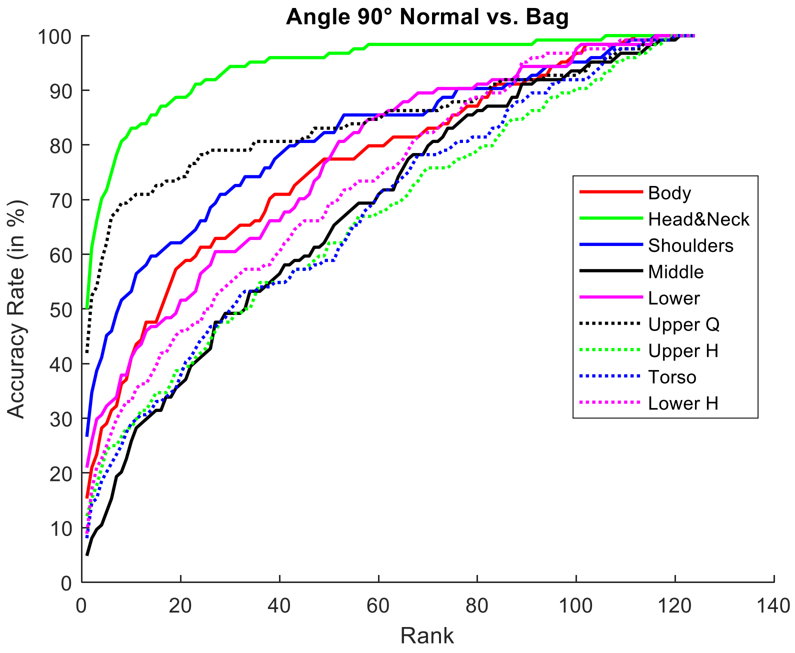


Figure 5.13: Normal versus Bag cross-scenario approach CMC curves based on implementation of DTW from Angle 90°

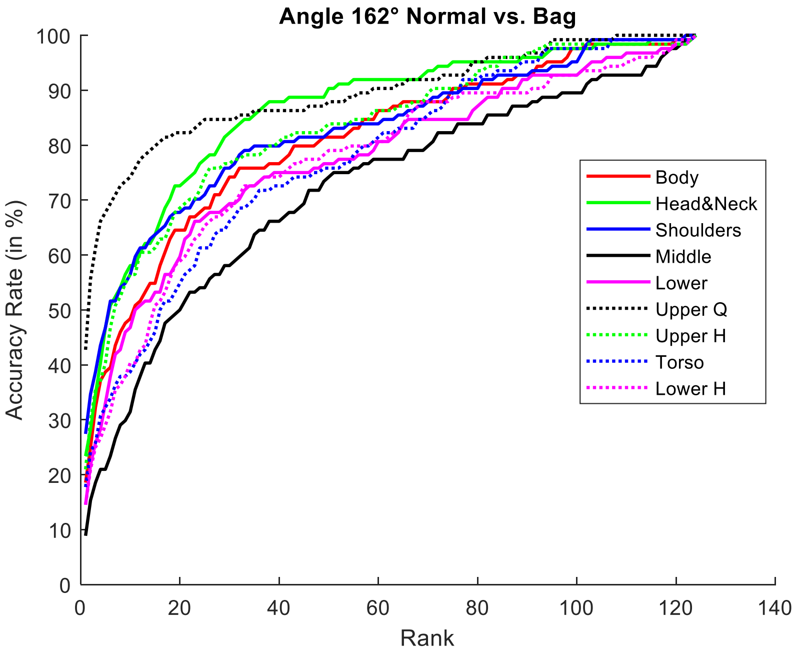


Figure 5.14: Normal versus Bag cross-scenario approach CMC curves based on implementation of DTW from Angle 162°

Table 5.3: Cross-scenario first rank rates (in percentage) of Normal versus Bag scenario

Segment Angle	Body	Head & Neck	Shoulders	Middle	Lower	Upper Q	Upper H	Torso	Lower H
0°	36	24	25	15	25	**49**	28	19	23
18°	17	16	19	14	12	**26**	12	16	16
36°	16	22	15	10	9	**24**	15	13	12
54°	15	**19**	11	7	15	**18**	12	10	9
72°	11	**37**	14	2	5	31	12	6	4
90°	15	**50**	26	4	20	41	12	8	8
108°	12	**36**	13	6	16	34	12	8	13
126°	11	**34**	11	6	10	27	11	8	8
144°	14	**32**	15	6	16	25	12	9	12
162°	18	23	27	8	14	**42**	20	17	14
180°	42	23	30	16	25	**50**	27	20	24

5.5.2 Normal versus Clothes Video Sequences

Experiments like those used for the Normal versus Bag scenario were performed for the Normal versus Clothes scenario to explore DTW performance in the case where the subject wears/takes off heavy clothes. The purpose of this experiment was to explore whether this feature (i.e., DTW) would be adequate for identifying the dynamic discrimination of the shape description.

This experiment was implemented using a methodology similar to that used for the previous scenario described. However, in this trial the DTW algorithm was applied to two shape description sequences, where the training sequence was in the Normal scenario and the test sequence was in the Clothes scenario.

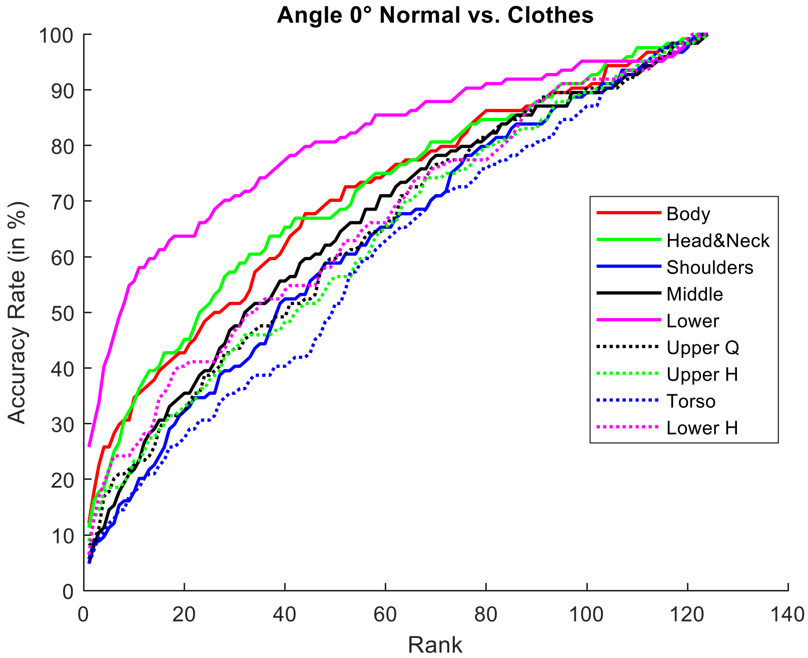


Figure 5.15: Normal versus Clothes cross-scenario approach CMC curves based on implementation of DTW from Angle 0°

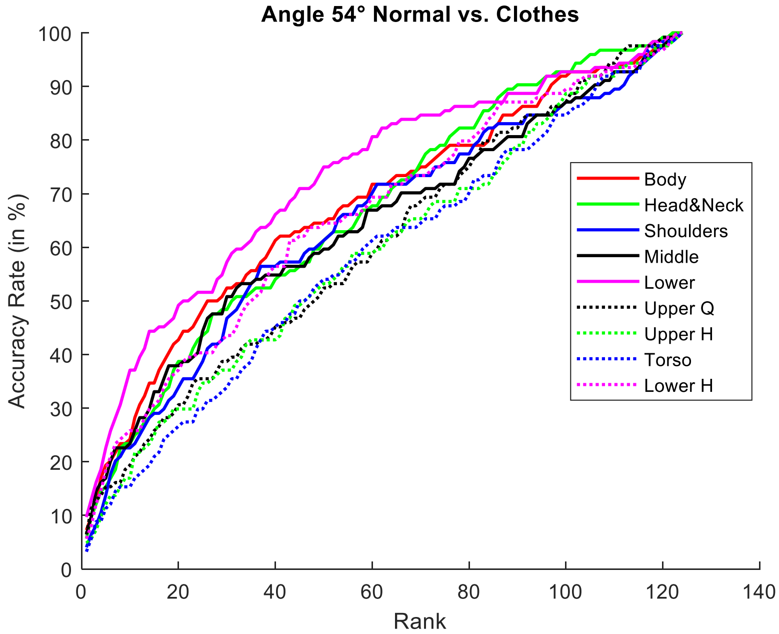


Figure 5.16: Normal versus Clothes cross-scenario approach CMC curves based on implementation of DTW from Angle 54°

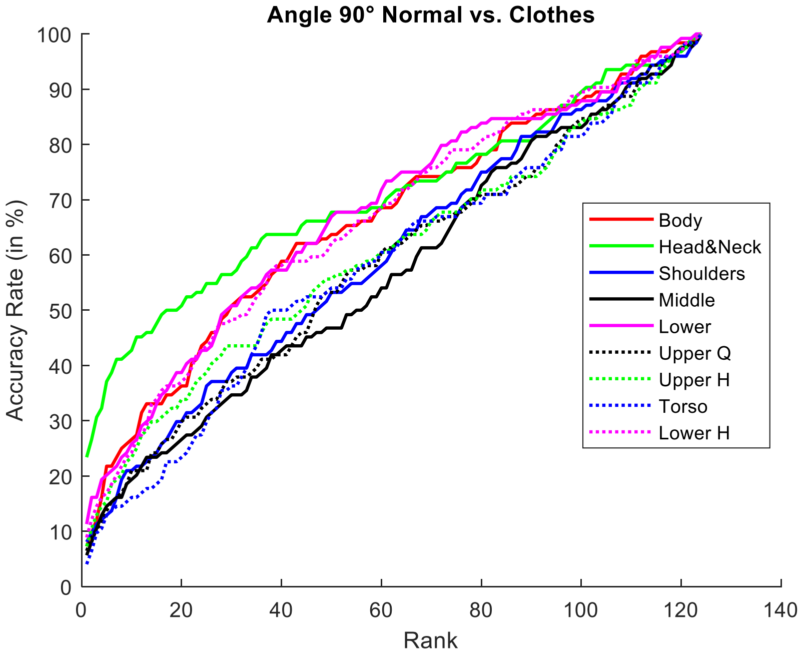


Figure 5.17: Normal versus Clothes cross-scenario approach CMC curves based on implementation of DTW from Angle 90°

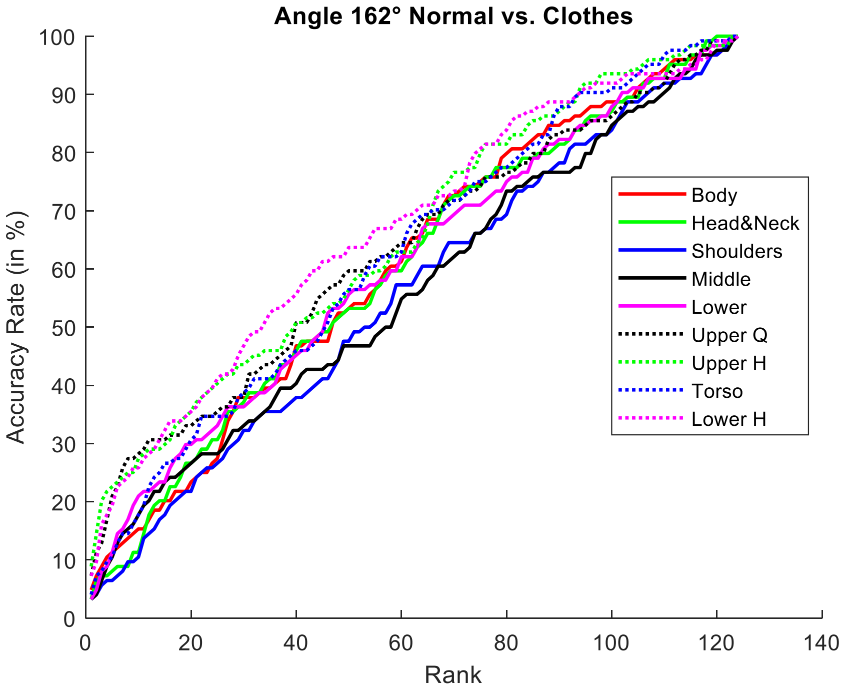


Figure 5.18: Normal versus Clothes cross-scenario approach CMC curves based on implementation of DTW from Angle 162°

Figure 5.15, Figure 5.16, Figure 5.17 and Figure 5.18 show the cross-scenario CMC curves of all proposed segments from four angles: 0°, 54°, 90°, and 162°, where the scenario was Normal versus Clothes.

Table 5.4 illustrates the tabular values of the first rank accuracy rates of all proposed segments from all viewing angles for the Normal versus Clothes scenario.

Results reveal that wearing heavy clothes affected a larger part of the body than carrying a bag, as it covered the shoulders, arms, and part of the leg. This can be observed in the CMC curves shown in Figure 5.9 and in the first rank accuracy rates provided in Table 5.4. All segments showed extremely low accuracy rates compared to the Normal versus Normal and Normal versus Bag results, except for the *Lower* segment, which remained in a similar range of accuracy rates.

Similar to the findings in the previous section, although the heavy clothes did not directly affect the shape of the *Lower* segment, the *Lower* segment accuracy rates in Normal versus clothes were lower than the results of the same segment in the Normal versus Normal approach. This proves that the heavy clothes influence the pose and the motion of the *Lower* segment, causing a variation in the *Lower* segment shape description.

5.6 Comparison with Related State-of-Art Methods

The focus of this research is the static and dynamic discrimination levels of the shape description of the body and body segments. The literature review show no evidence of experimental study has been published in which a shape descriptor is used on a segmented human silhouette. As the work presented here is based on one feature (GFD) and followed with one classification method, either LDA for image-based system or DTW for video-based system. It was difficult to draw a fair and direct comparison with approaches that utilised multiple features and different fusion levels. Therefore, the comparisons here are between the results of the *Body* segment and some related state-of-art person re-id methods that were tested on the CASIA Dataset B. The values of our method presented in the tables in this section show the average first rank accuracy rates of all 11 viewing angles. Hence, this section provides a comparison of the general performance using the shape descriptor on the entire body silhouette as an identifier for person re-id applications.

Table 5.4: Cross-scenario first rank accuracy rates (in percentage) of Normal versus Clothes scenario

Segment **Angle**	**Body**	**Head & Neck**	**Shoulders**	**Middle**	**Lower**	**Upper Q**	**Upper H**	**Torso**	**Lower H**
0°	12	11	5	4	**25**	8	8	6	6
18°	**13**	8	4	6	10	6	5	4	5
36°	10	8	8	8	**11**	**11**	6	7	7
54°	7	5	4	6	9	6	4	3	5
72°	7	10	5	6	**11**	6	4	6	7
90°	7	**23**	8	5	11	6	7	4	8
108°	9	**14**	7	2	10	7	4	5	10
126°	10	**16**	6	5	13	6	4	4	10
144°	8	8	6	6	**12**	6	7	5	8
162°	4	3	3	3	3	7	**8**	4	7
180°	16	8	4	8	**29**	6	7	4	6

Method 1 [74] enhanced the appearance based person re-id using HSV colour feature and Gabor texture feature by integrating the gait feature. The features were fused in two different ways, namely, feature-level fusion and score-level fusion.

Method 2 [75] involved dividing the human silhouette into three parts: head, torso, and leg segments. Each segment was then described using the HSV colour feature and the maximally stable colour region (MSCR).

Method 3 [91] used AdaBoost learning on the same features used in *Method 2*.

Method 4 [138] and *Method 5* [139] used ITML and LMNN metric learning methods, respectively, on the same features. [138] and [139] divided the person silhouette into six horizontal segments. They then extracted the RGB, YCbrCr, and HSV colour features and the Gabor and Schmid texture features as well.

5.6.1 Angle-based Approach Comparison

In this section, the results of implementing LDA and DTW on the *Body* segment shape description is compared with the related state-of-arts methods (1–5) referenced. The approach of these studies (including our methods) is angle-based comparison, where the training and test sequences were captured from the same viewing angle. In addition, the two sequences are within the same scenario, which is Normal versus Normal.

Table 5.5 reports overall performance data for the different state-of-arts approaches along with performance data for the proposed methods. The results indicate the potentially superior performance of the proposed approaches, especially at the first 10 ranks.

Table 5.5: State-of-arts top ranked accuracy rates (in percent) of person re-id for Normal versus Normal scenario

Method	$r = 1$	$r = 5$	$r = 10$	$r = 15$	$r = 20$
Method 1 (feature-level fusion)	16.29	43.44	60.75	72.17	79.40
Method 1 (score-level fusion)	13.55	48.74	63.73	72.76	79.55
Method 2	4.90	27.04	41.55	52.28	60.49
Method 3	12.25	35.55	50.25	60.17	66.87
Method 4	7.48	22.21	34.15	43.49	50.07
Method 5	3.89	22.65	36.06	46.41	54.32
Our method (GFD+LDA)	**49.85**	**74.67**	**82.98**	**87.18**	**89.83**
Our method (GFD+DTW)	**46.77**	**63.27**	**70.60**	**75.73**	**79.84**

5.6.2 Cross-scenario Approach Comparison

Further comparison was conducted to examine the performance of the cross-scenario methodology implemented by the same studies. Table 5.6 depicts our experimental results along with the results of other methods based on the cross-scenario, i.e., the Normal versus Bag scenario and Normal versus Clothes scenario. Although Method 1 (feature-level fusion) outperformed the proposed methods, the proposed methods still presented results akin to those of the other studies.

Table 5.6: State-of-art top ranked accuracy rates (percent) of person re-id on Normal versus Bag scenario

Method	$r = 1$	$r = 5$	$r = 10$	$r = 15$	$r = 20$
Method 1 (feature-level fusion)	**31.81**	**53.59**	**64.14**	**70.48**	**77.03**
Method 1 (score-level fusion)	14.67	32.59	50.16	60.55	67.29
Method 2	22.91	30.09	36.07	41.09	48.03
Method 3	17.14	30.01	37.85	44.33	52.90
Method 4	21.83	30.41	36.28	41.27	48.04
Method 5	23.11	37.11	44.44	49.38	56.85
Our method (GFD+LDA)	22.46	44.81	56.71	64.26	69.72
Our method (GFD+DTW)	19.28	35.97	43.62	50.73	56.89

Table 5.7 provides the proposed accuracy rates of the Normal versus Clothes scenario along with the rates of other methods. Method 1 (feature-level fusion) outperformed in this approach as well; however, the rates of the proposed method fell within the range of the rates of the other compared methods.

Table 5.7: State-of-art top ranked accuracy rates (percent) of person re-id on Normal versus Clothes scenario

Method	$r = 1$	$r = 5$	$r = 10$	$r = 15$	$r = 20$
Method 1 (feature-level fusion)	**20.28**	**42.64**	**56.87**	**67.81**	**75.02**
Method 1 (score-level fusion)	9.68	27.77	45.12	54.23	60.67
Method 2	11.64	19.38	27.57	35.71	42.10
Method 3	5.63	15.99	26.34	36.63	45.21
Method 4	10.27	24.48	36.11	47.03	55.40
Method 5	11.61	12.62	17.75	24.22	29.35
Our method (GFD+LDA)	10.47	25.22	35.18	42.41	48.37
Our method (GFD+DTW)	9.82	20.67	27.49	34.02	38.93

Carrying a bag and wearing heavy clothes directly affect the body shape of the subjects which prevents the body shape descriptor in cross-scenario from outperforming the combination of colour, texture and gait feature presented in [74]

5.7 Summary

This chapter describes the implementation of DTW on the GFD shape description for person re-id application. The proposed system employs the dynamic feature of the shape description to find correct identity matches. The main contributions presented within this chapter are as follows:

- Adapting the one-dimension DTW to be performed on the multidimensional data of the examined dataset.
- Performing the multidimensional DTW algorithm on the proposed body and body segments shape descriptor.
- Analysing the performance of the proposed body segments to identify direct and indirect factors that negatively affect performance.
- Implementing the multidimensional DTW algorithm on inter-view sequences.
- Discovering the performance of the multidimensional DTW algorithm on two different scenarios, i.e., Normal versus Bag and Normal versus Clothes.
- Comparing the performance of the proposed systems on several related state-of-art person re-id studies.

The main conclusions can be summarised as follows:

- The implementation of DTW revealed unique dynamic features between the subjects' GFD shape descriptions.
- The dynamic features of the body and body segments' performance varied from one angle to another.
- Some segments performed better in the straight views, while others performed better in the side views.
- Implementing DTW on the GFD feature vector (video-based approach) outperformed implementation of the LDA on the GFD feature vector (image-based).

- The dynamic feature on the inter-view and cross-scenario approaches presented a low level of accuracy compared to the original scenario (angle-based and normal versus normal approach).

In the next Chapter, the performance of the two proposed systems is compared, and a number of rank list fusions are conducted in pursuit of additional performance improvements for all discussed approaches.

Chapter 6

Systems Analysis and Rank Lists Fusion

6.1 Introduction

In Chapters 3, 4, and 5, comprehensive examinations of the use of body and body segment shape descriptors as identifiers for person re-identification (re-id) were presented. This Chapter complements those examinations through further analysis of the data gleaned thus far, using three approaches.

First, the overall outputs of the person re-id systems are compared and analysed. More specifically, the analysis focuses on the performances of the proposed image-based and video-based person re-id systems by assessing and comparing their accuracy rates. This comparison is intended to identify the general performance of and highlight the outperforming segment for each system, thus pinpointing the most trusted part of the body (i.e., segment) for use of the image-based and video-based person re-id systems.

Second, the output rank lists of the different systems are compared. The rank lists differ for the image-based and video-based systems. The image-based system generates one rank list for each frame, whereas the video-based system generates one rank list for the entire sequence (each sequence consisting of multiple frames). In order to apply fair systems comparison, the number of the generated rank lists of the image-based system is necessary to be equal to the number of the generated video-based rank lists. This is accomplished by fusing the generated rank lists of the image-based system, into one rank list for multiple frames (i.e. sequence). This is to parallel the structure of the video-based system rank lists, which enables the comparison of the accuracy rates of the two systems.

The third analysis approach described in this chapter involves exploiting the image-based rank lists fusion approach to potentially identify performance improvements. This method is used because in the inter-view experiment conducted in Section 4.3, performance was low compared to the angle-based experiment in Section 4.4.

Similarly, the accuracy rates were considerably low for the cross-scenario experiment presented in Section 4.5, compared with the rates of the angle-based experiment discussed in Section 4.3. Thus, the image-based rank lists fusion approach for cross-scenario performance improvements will also be examined.

Figure 6.1 shows the general framework of the shape-based person re-id system, highlighting the content of the current Chapter, which primarily includes data on the analysis and comparison of system outputs and on the implementation of rank lists fusion on several aspects of the research.

This Chapter is organised as follows: Section 6.2 compares the general performances of image-based and video-based systems and discusses the outputs of these comparisons. Section 6.3 proposes a rank lists fusion approach, which is implemented on the image-based rank lists in Section 4.3, on the inter-views in Section 4.4, and on cross-scenario approaches in Sections 4.5.1 and 4.5.2. Finally, Section 6.4 summarises the main findings of the system analyses and rank lists fusions.

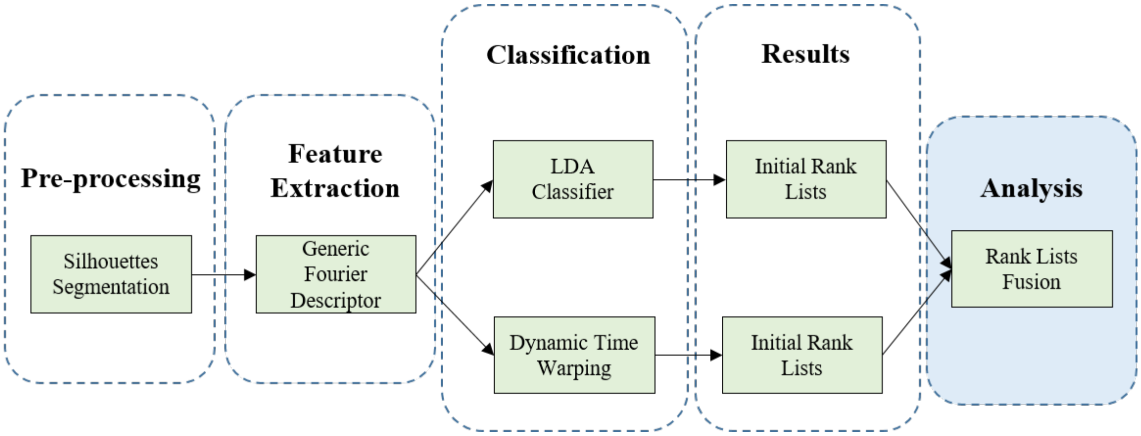


Figure 6.1: Proposed shape-based person re-id system framework, highlighting the focus of the current Chapter.

6.2 Segment-based Systems Analysis

To consider the baseline aspects of shape-based person re-id, the performances of the proposed image-based and video-based systems were compared. The comparisons were based on the performances of both systems on the same body segment. In this research, the Generic Fourier Descriptor (GFD) shape descriptor was extracted from each frame, forming an individual feature vector. In the image-based system, the Linear Discriminant Analysis (LDA) classifier was implemented on the feature vectors, resulting in an individual rank list for each frame, while in the video-based system, a group of feature vectors of all frames comprising one sequence was used, as one sequence represented one subject. The Dynamic Time Wrapping (DTW) algorithm was applied to the entire sequence of feature vectors to be aligned with all sequences of the test set. Finally, one rank list was generated for each sequence. The rank list contained the subjects' identities and was ordered based on the DTW distance with the current examined sequence.

Therefore, the first rank of the image-based system represented the number of the correct subject identity matches between all the frames in the training set and test set, whereas the first rank of the video-based system represented the number of the correct subject identity matches between the sequences of the training set and test set. Presented in the figures below are the first rank for each segment for all viewing angles from both systems. In the next subsections, segments are discussed individually.

Body;

Figure 6.2 shows the first rank data for the *Body* segment in all the angles. The accuracy rates of both systems were extremely close from each angle for this segment. The image-based performance noticeably increased in the side view and the angles that primarily capture side views, such as 72°, 90°, and 108° angles. The average differences in performance between both systems were 10%, 17%, and 8%, respectively. Also, there was a noticeable increase in the video-based system performance for the straight views (0° and 180°).

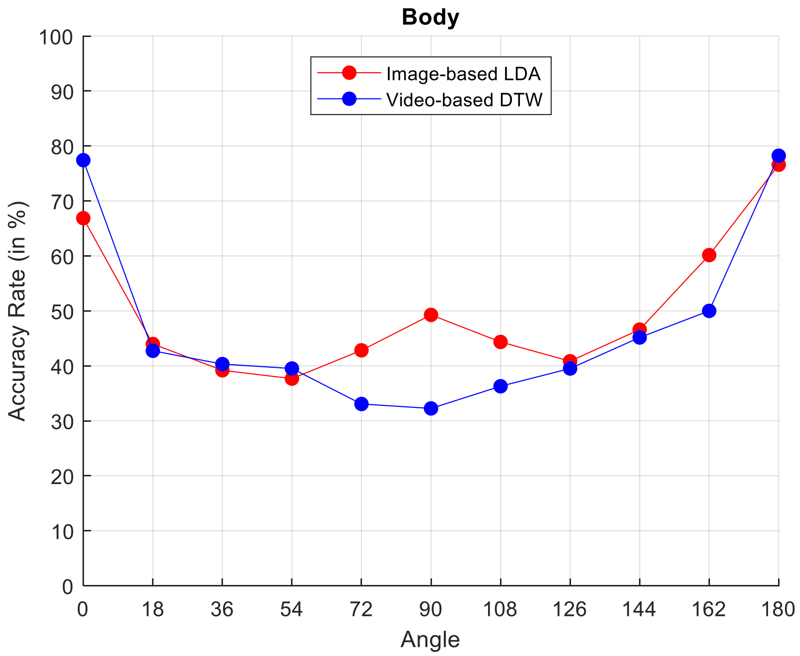


Figure 6.2: First rank accuracy rates of the *Body* segment from all 11 viewing angles in both proposed systems using image-based LDA classifier and video-based DTW algorithm.

Head & Neck; the video-based system outperformed the image-based system for the first seven angles for the *Head & Neck* segment, but their performances were almost equal for the last four angles. This is shown in Figure 6.3.

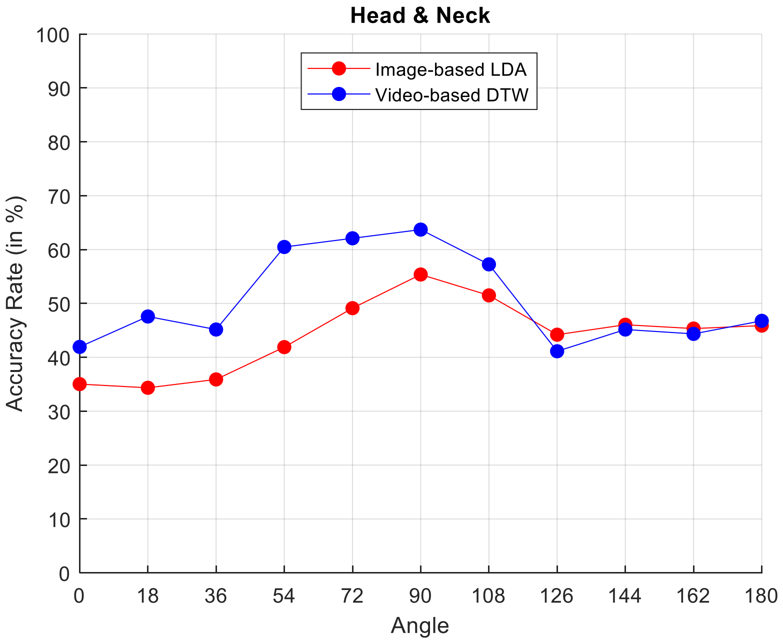


Figure 6.3: First rank accuracy rates of the *Head & Neck* segment from all 11 viewing angles in both proposed systems using image-based LDA classifier and video-based DTW algorithm.

Shoulders; although the systems did not reflect a pattern in performance between the viewing angles of the *Shoulders* segment, Figure 6.4 shows the significant performance increase in the video-based system compared to the image-based performance.

Middle; the performance of the *Middle* segment was, on average, 20% higher when using the DTW algorithm (video-based system) than when using the LDA classifier (image-based system). Figure 6.5 shows the significant performance enhancement.

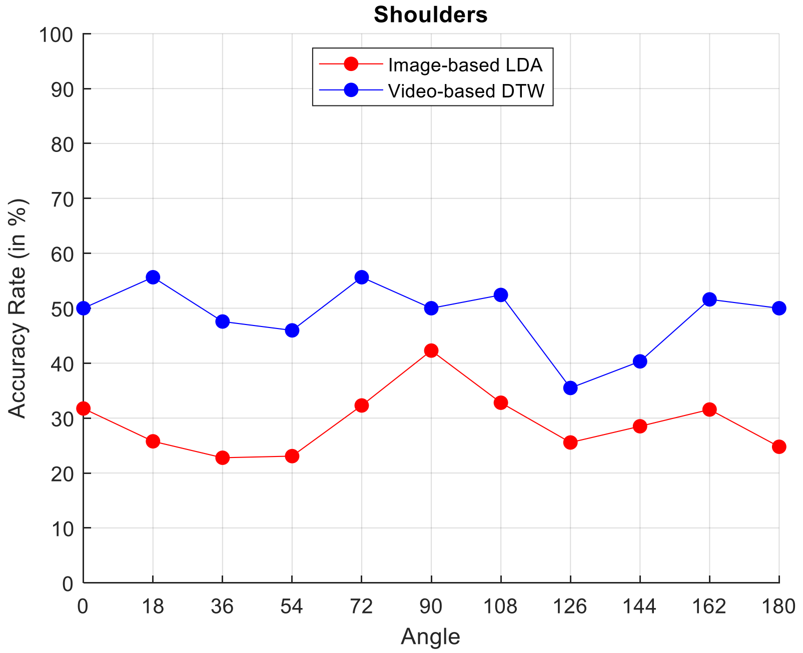


Figure 6.4: First rank accuracy rates of the *Shoulders* segment from all 11 viewing angles in both proposed systems using image-based LDA classifier and video-based DTW algorithm.

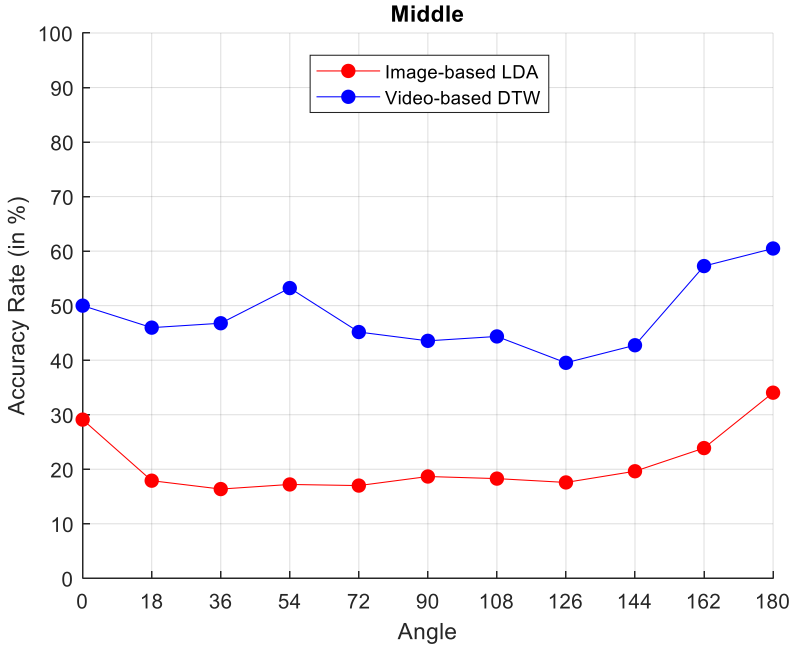


Figure 6.5: First rank accuracy rates of the *Middle* segment from all 11 viewing angles in both proposed systems using image-based LDA classifier and video-based DTW algorithm.

Lower; DTW algorithm implementations on the *Lower* segments outperformed the implementation of the LDA. The DTW algorithm achieved 15% higher first rank accuracy rates than the LDA. Figure 6.6 shows the results gained from both systems.

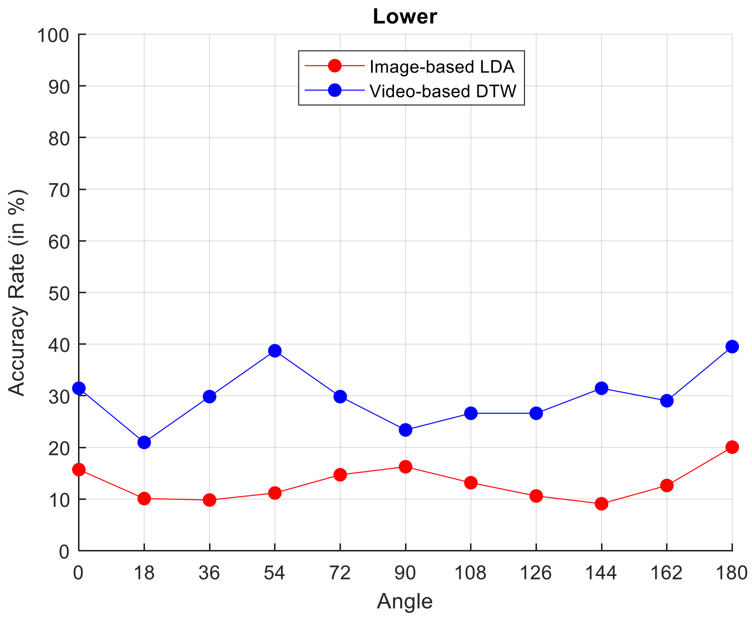


Figure 6.6: First rank accuracy rates of the *Lower* segment from all 11 viewing angles in both proposed systems using image-based LDA classifier and video-based DTW algorithm.

Upper Q; this is one of the outperforming segments in both the image-based and video-based systems. Implementation of DTW showed noticeable performance development compared with use of the LDA. Figure 6.7 shows the accuracy rate differences between the image-based and video-based systems for the *Upper Q* segment.

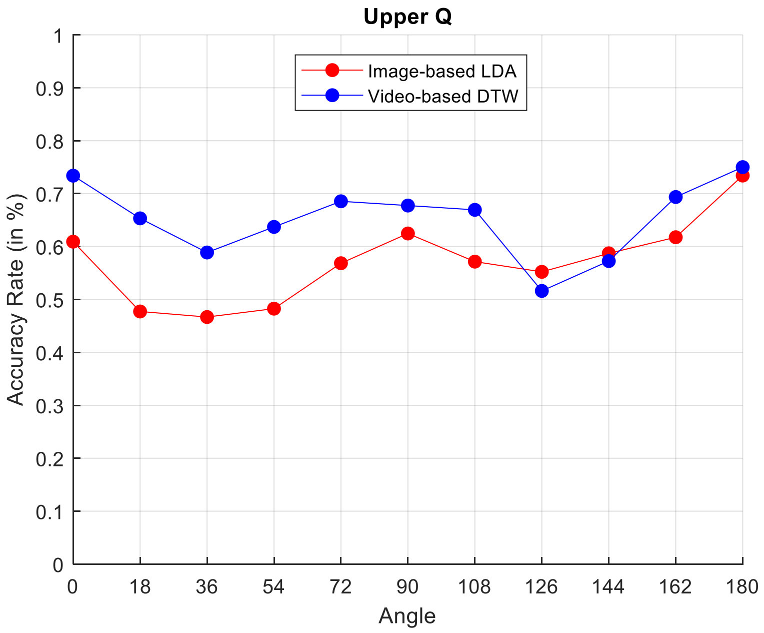


Figure 6.7: First rank accuracy rates of the *Upper Q* segment from all 11 viewing angles in both proposed systems using image-based LDA classifier and video-based DTW algorithm.

Upper H; this is another segment with high performance in both systems. The general performance of the image-based and video-based systems are similar for this segment. Figure 6.8 compares the performance data of the *Upper H* segment resulting from the application of the LDA classifier and DTW algorithm.

Torso; following the *Middle* and *Lower* segments, DTW implementation on the *Torso* segment produced outperforming accuracy rates compared with the accuracy rates for the application of the LDA classifier. Figure 6.9 shows performance for both systems.

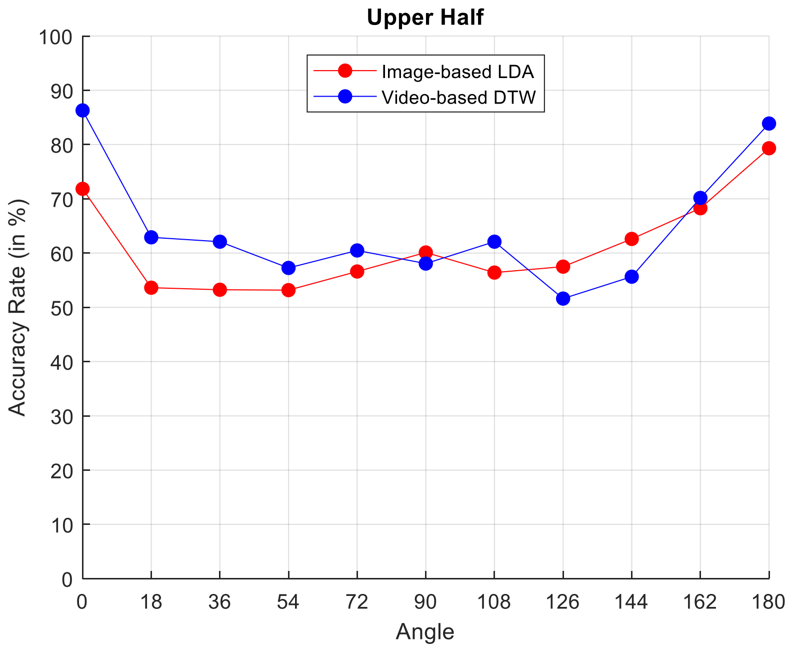


Figure 6.8: First rank accuracy rates of the *Upper H* segment from all 11 viewing angles in both proposed systems using image-based LDA classifier and video-based DTW algorithm.

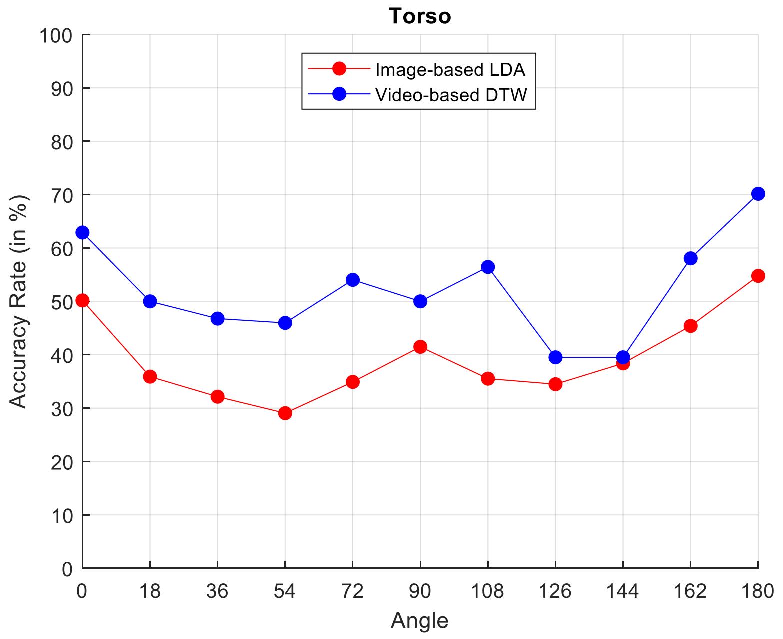


Figure 6.9: First rank accuracy rates of the *Torso* segment from all 11 viewing angles in both proposed systems using image-based LDA classifier and video-based DTW algorithm.

Lower H; this segment contains two of the most movable parts of the body—lower arms with hands and leg parts. Therefore, the accuracy rates generated by the DTW algorithm outperformed the rates generated by the LDA classifier, which classify the shape description feature vectors frame-by-frame, while the DTW algorithm matches sequences based on their dynamic features. Figure 6.10 shows the data from both systems.

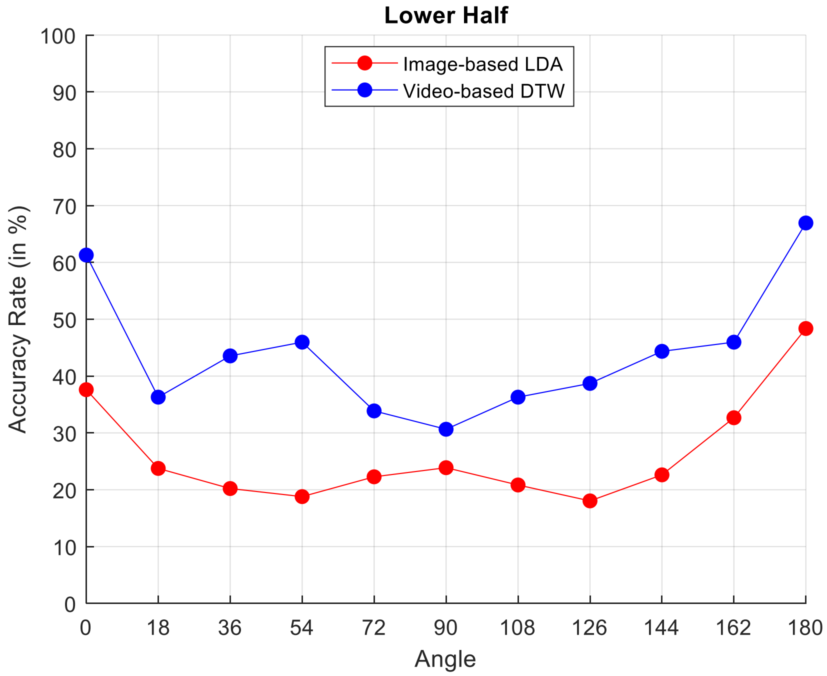


Figure 6.10: First rank accuracy rates of the *Lower H* segment from all 11 viewing angles in both proposed systems using image-based LDA classifier and video-based DTW algorithm.

To summarise, the findings of the segment-based system comparisons presented in this section demonstrate that greater accuracy rates were achieved through implementation of the DTW algorithm on video-based images than from applying the LDA classifier to the image-based system. This proves that the dynamic variation of shape descriptors is more discriminative for person re-id than the static variation. In Figure 6.11, the accuracy rates for each segment at all angles are averaged, revealing the overall performance for each segment from both the image-based and video-based systems.

Another important finding from this analysis is that comparing the performance of both systems on the segments that included hands, arms and/or legs, such as the *Middle, Torso, Lower,* and

Lower H segments, shows that significant performance improvements are achieved when using the DTW algorithm to find unique dynamic features between the subjects' shape descriptions.

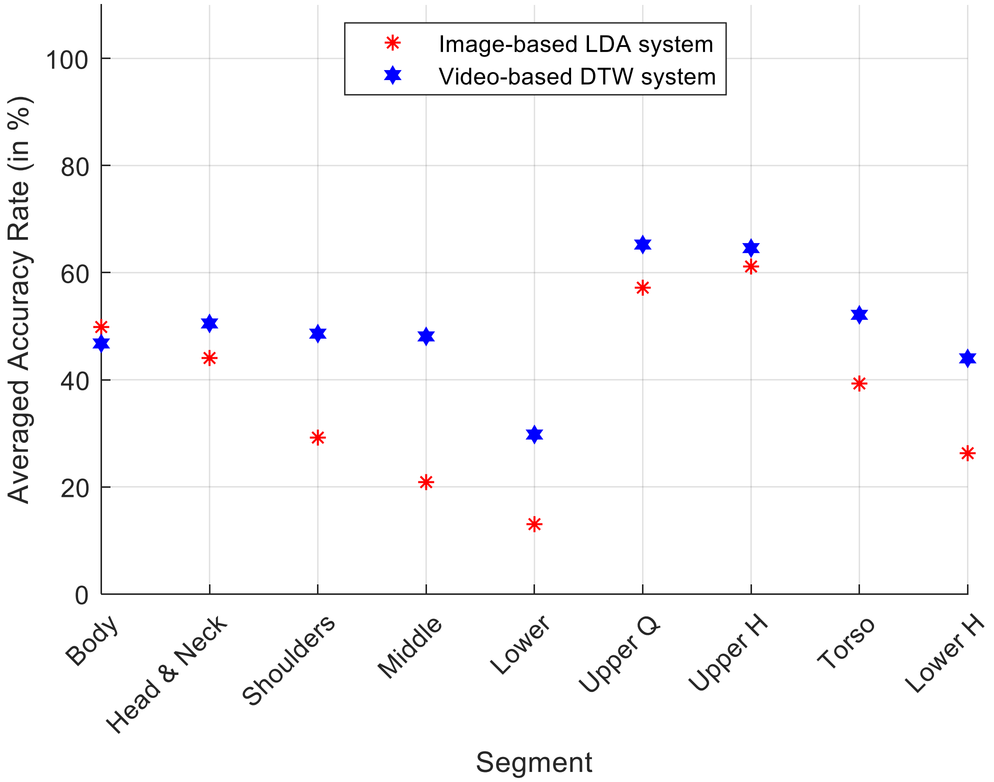


Figure 6.11: The averaged accuracy rates for each segment at all angles from image-based and video-based systems.

6.3 Rank Lists Fusion

In person identification, rank level fusion is the process of combining more than one identification results in order to improve the performance. Rank level fusion is a method applied on multimodal biometric systems, which combines the scores of different biometric systems (i.e. face, fingerprint and iris). In addition, rank level fusion can combine the results of multiple classifier, training set or parameter values of one biometric model.

There are number of rank level fusion methods found in the literature. For example, Borda count approach [140], which based on the generalization of majority vote and the most commonly used approach for unsupervised rank-level fusion [141].

Our image-based system was designed to implement the frame-by-frame classification process. This means that each frame was generated with an individual rank list that ordered the current training list (i.e., subjects), ranking them from most similar to least similar. These multiple rank lists for each subject's sequence (or multiple frames) provided an opportunity for performance enhancement by exploiting the lists.

As part of this research, the multiple rank lists obtained from the image-based system of each sequence were fused to generate one rank list for each sequence with the goal of system performance enhancement using Borda count approach. This rank lists fusion was implemented on the original image-based rank lists, inter-views, and cross-scenario approaches. These are explained and discussed in the next subsections.

6.3.1 Image-based Fusion

As previously discussed, the image-based system generates multiple rank lists for each sequence (on a frame-by-frame basis). Borda count rank lists fusion combined these lists so that the set of lists generated by all frames in each sequence was replaced by one rank list for each sequence. The main purpose using this rank list approach was to count the indices of each identity in the initial rank lists. The indices of each identity from each initial rank list were added together, and then the identity with the least total indices was placed in the frontal location of the new fused rank list. This can be formulated as follows: Considering the sequence of multiple rank lists is:

$$S = \begin{bmatrix} RL_1 \\ RL_2 \\ \vdots \\ RL_x \end{bmatrix} \tag{6.1}$$

where RL represents the rank list for each frame in the sequence S, and x is the number of rank lists (or frames) in S. RL can be formulated as follows:

$$RL = [r_1, r_2, \dots, r_n] \tag{6.2}$$

where the r represents the identities that were initially ordered based on their similarity to the description of the current frame, and n is the length of the watch list or the number of subjects in the dataset.

Table 6.1 illustrates the technical steps for implementing the rank lists fusion.

Table 6.1: The pseudocode of the rank lists fusion approach using Borda count approach.

```
Rank Lists Fusion Algorithm
Input: $S$
Output: $fusedRL$

Let $identities$ be the matrix that contains all subjects' identities in the dataset along with their total indices in all the rank lists of $S$
for $i = 1$ to $x$ do
    for $j = 1$ to $n$ do
        1. Add $RL_i(r_j)$ to $identities$ if it does not already exist
        2. Add $j$ to the total indices of this identity
    end for
end for
$fusedRL$ = sort $identities$ (in ascending order based on their total indices)
```

This method was implemented on the rank lists generated by the image-based system discussed in Section 4.3. The performance for the fused rank lists is presented in two stages. First, the initial rank lists accuracies for each segment from all angles were compared with the accuracy rates of the fused rank list. These comparisons are shown in Table 6.2.

The other procedure for evaluating the rank lists fusion was to compare it with the performance obtained from the video-based systems. This comparison is presented in Figure 2.12, where the performances at all angles are averaged for each segment. The performance from original image-based, video-based, and image-based fused rank list were then compared.

Table 6.2: The first rank accuracy rates of the initial image-based and the fused rank list (in percentage), with the outperforming rank list for each segment and angle highlighted.

Angle **Segment**	**0°**	**18°**	**36°**	**54°**	**72°**	**90°**	**108°**	**126°**	**144°**	**162°**	**180°**
Body	66	43	39	37	42	49	44	40	46	60	76
Body fused	**95**	**91**	**95**	**93**	**96**	**94**	**93**	**96**	**95**	**95**	**99**
Head & Neck	35	34	35	41	49	55	51	44	46	45	45
Head & Neck fused	**86**	**88**	**85**	**91**	**89**	**94**	**91**	**91**	**87**	**96**	**92**
Shoulders	31	25	22	23	32	42	32	25	28	31	24
Shoulders fused	**83**	**73**	**73**	**84**	**86**	**84**	**82**	**83**	**85**	**84**	**66**
Middle	29	17	16	17	17	18	18	17	19	23	34
Middle fused	**83**	**69**	**77**	**79**	**67**	**69**	**76**	**79**	**83**	**87**	**96**
Lower	15	10	9	11	14	16	13	10	9	12	20
Lower fused	**67**	**51**	**54**	**61**	**71**	**74**	**68**	**65**	**52**	**71**	**78**
Upper Q	60	47	46	48	56	62	57	55	58	61	73
Upper Q fused	**97**	**95**	**95**	**96**	**96**	**94**	**91**	**95**	**96**	**95**	**97**
Upper H	71	53	53	53	56	60	56	57	62	68	79
Upper H fused	**97**	**95**	**95**	**95**	**95**	**98**	**93**	**95**	**95**	**97**	**99**
Torso	50	35	32	29	34	41	35	34	38	45	54
Torso fused	**95**	**89**	**91**	**94**	**93**	**95**	**91**	**95**	**94**	**92**	**96**
Lower H	37	23	20	18	22	23	20	18	22	32	48
Lower H fused	**87**	**87**	**87**	**76**	**84**	**82**	**80**	**81**	**91**	**95**	**96**

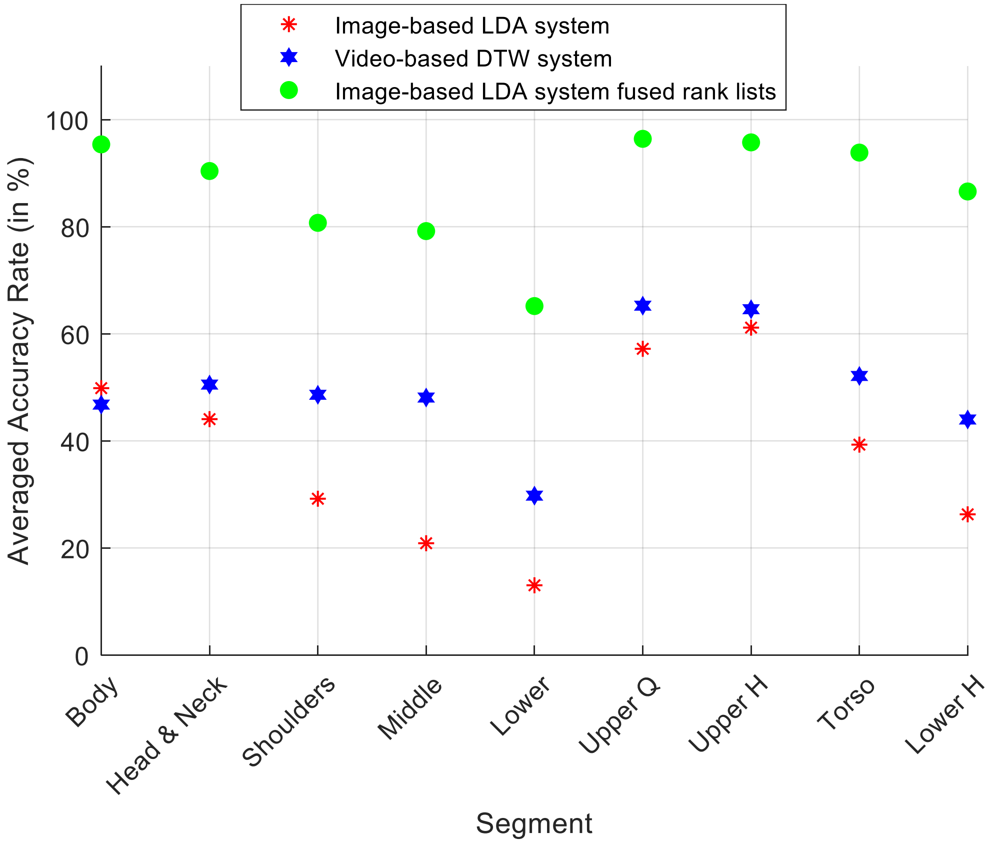


Figure 6.12: All angles average accuracy rates for each segment in three systems: image-based, video-based, and fused rank list image-based systems.

The results presented in Table 6.2 and Figure 6.12 show that rank lists fusion introduced significant performance enhancements to the shape-based person re-id. The image-based fused rank lists system outperformed both the original image-based system and video-based system. The main reason behind this vast improvement is that in the original systems, the rank list represents one source of data, which is one rank list for either a frame or a sequence. However, in the image-based fused rank lists system, the rank list represents multiple data sources, which are the multiple rank lists fused together for better identity estimation.

6.3.2 Inter-view Video Sequences Rank Lists Fusion

One of the expected scenarios in a real person re-id system is the identification of a subject captured from one viewing angle and the re-identification of the same subject captured from a different viewing angle. This concept, called *inter-views video sequences* in this research, was explored in this study within the two systems (i.e., image-based and video-based). The general performances when implementing the proposed systems in such scenarios were considerably low compared with the performances of same angle re-id.

In this phase of the study, the rank lists fusion approach that was discussed in Section 6.3.1 was exploited to enhance the performance of the inter-view video sequences shape-based person re-id. Technically, the rank lists obtained from the inter-view scenario of each sequence were fused following the process outlined in

Table 6.1. This generated one fused rank list for each sequence instead of multiple rank lists. The results of this implementation are presented along with the results from the original inter-views image-based system in Section 4.4 in order to determine if using rank lists fusion resulted in performance enhancement.

This is implemented on the rank lists obtained from the image-based system only because the image-based system produces multiple rank lists for each frame, allowing for rank lists fusion. It is not implemented on the video-based system, as the video-based system generates one rank list for each sequence, leaving no opportunity for further fusion.

The results of this examination are presented in a coloured map form for each segment in Figure 6.13–Figure 6.21. Analysing these results revealed significant improvements in the performance of the inter-view approach. This confirms the positive influence of fusing the rank lists on the shape-based person re-id.

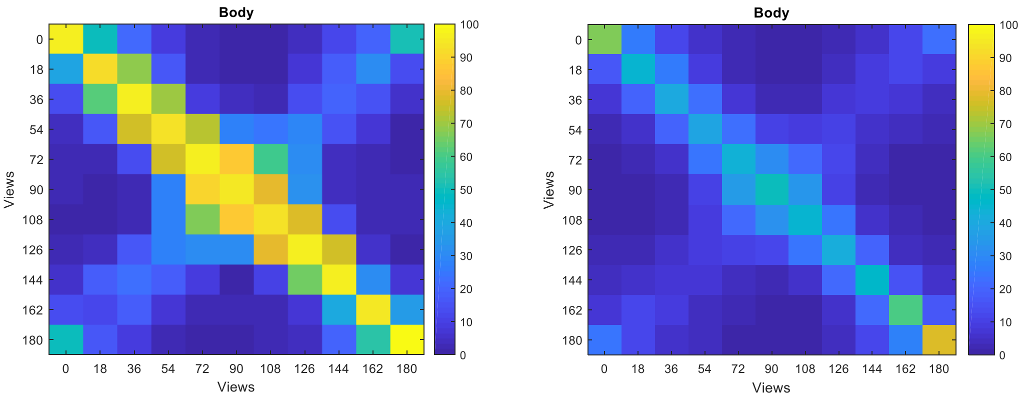


Figure 6.13: Right—the original inter-view *Body* segment performance; Left—the rank lists fusion inter-view *Body* segment performance.

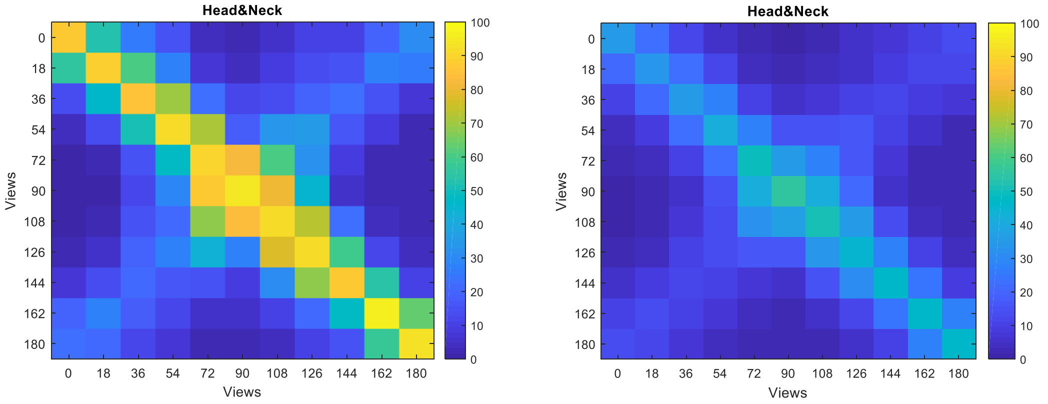


Figure 6.14: Right—the original inter-view *Head & Neck* segment performance; Left—the rank lists fusion inter-view *Head & Neck* segment performance.

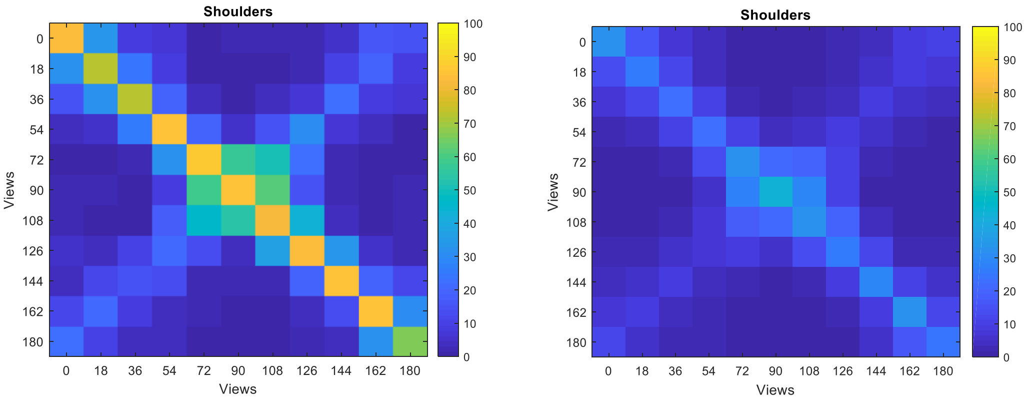


Figure 6.15: Right—the original inter-view *Shoulders* segment performance; Left—the rank lists fusion inter-view *Shoulders* segment performance.

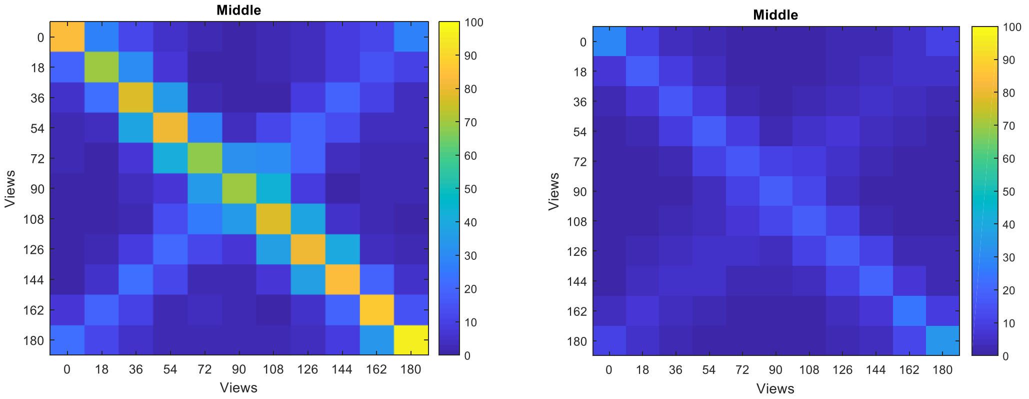


Figure 6.16: Right—the original inter-view *Middle* segment performance; Left—the rank lists fusion inter-view *Middle* segment performance.

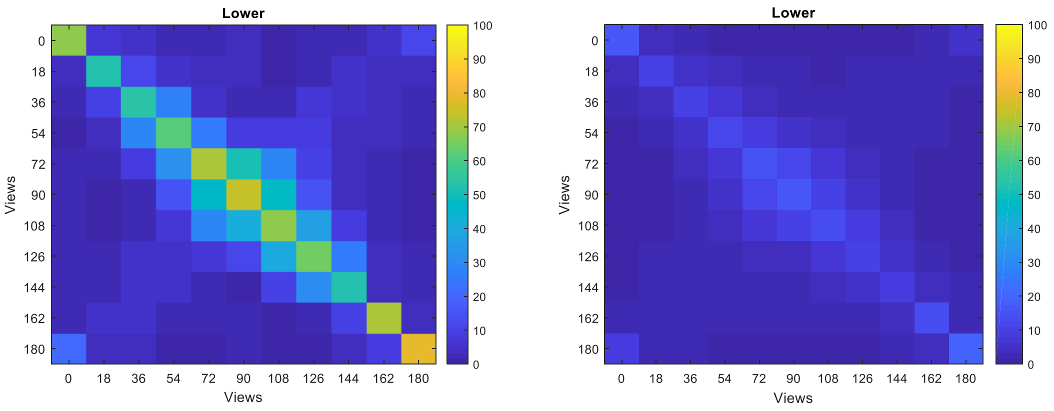


Figure 6.17: Right—the original inter-view *Lower* segment performance; Left—the rank lists fusion inter-view *Lower* segment performance.

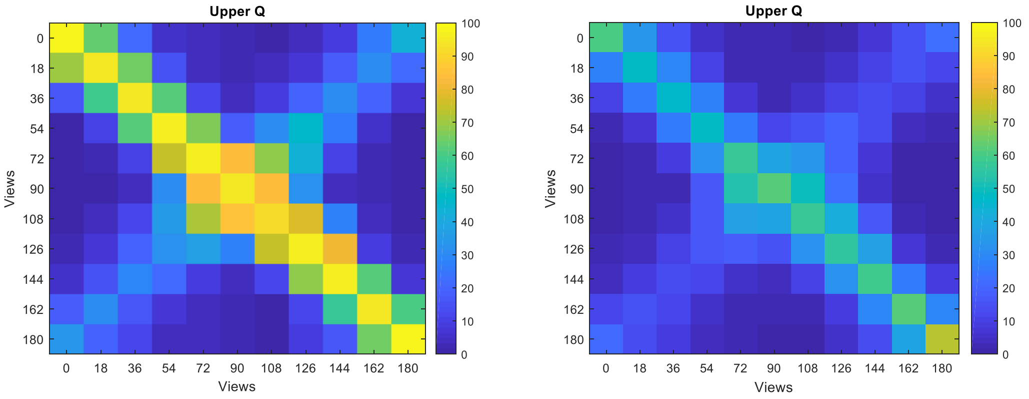


Figure 6.18: Right—the original inter-view *Upper Q* segment performance; Left—the rank lists fusion inter-view *Upper Q* segment performance.

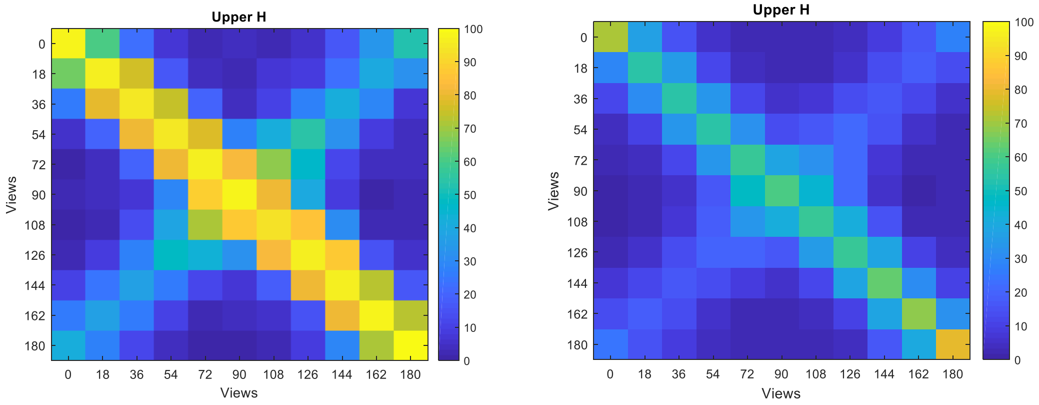


Figure 6.19: Right—the original inter-view *Upper H* segment performance; Left—the rank lists fusion inter-view *Upper H* segment performance.

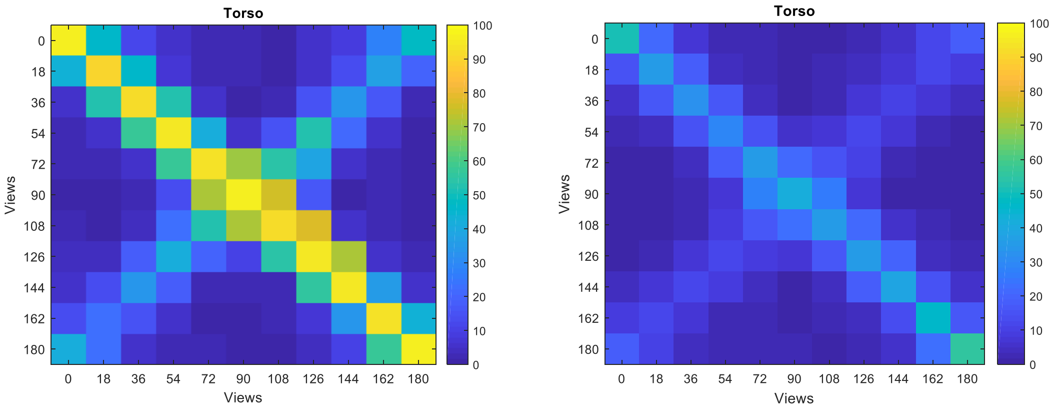


Figure 6.20: Right—the original inter-view *Torso* segment performance; Left—the rank lists fusion inter-view *Torso* segment performance.

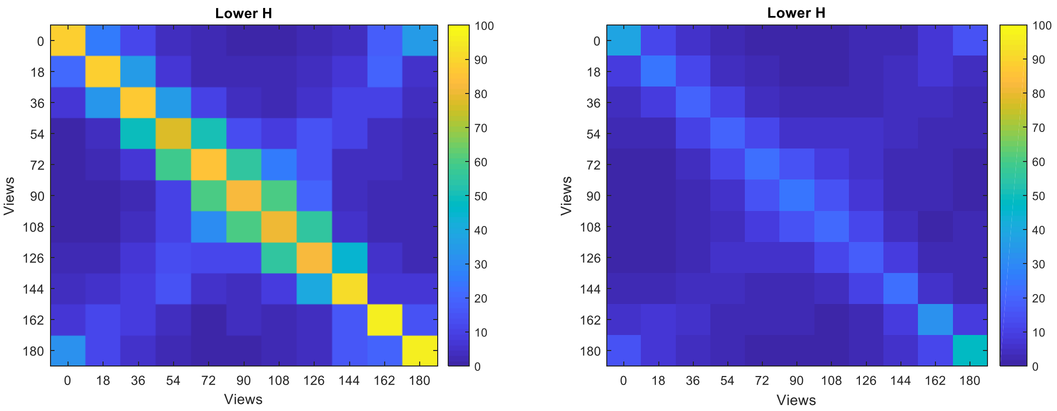


Figure 6.21: Right—the original inter-view *Lower H* segment performance. Left; the rank lists fusion inter-view *Lower H* segment performance.

6.3.3 Rank Lists Fusion for Cross-Scenario Performance Improvements (Normal vs. Bag)

The re-id performance using two different appearances, such as Normal and Carrying a Bag, was considerably low compared to the re-id performance using the same appearance (i.e., Normal vs. Normal). As with previous aspects of the study presented in this chapter, further rank lists fusion experiments were conducted on the cross-scenario seeking performance improvements. Technically, the rank lists obtained from the cross-scenario image-based system in Section 4.5 were fused utilising the rank lists fusion explained in

Table 6.1 The accuracy rates of the first rank of the original results along with rank lists fusion results are presented in Table 6.3.

The results presented in Table 6.3 reflect significant improvements in the performance when identifying a subject in normal appearance and re-identifying them wearing a bag. In this table, the outperformed scheme was highlighted. For all segments, rank lists fusion outperformed the original system.

Although the *Middle* and *Torso* segments showed performance improvements, their improvement levels were minimal compared with those of other segments. The primary reason for this is that the bags mostly affected the shape of these segments (i.e., *Middle* and *Torso*), as they were mostly located in those parts of the body.

Table 6.3: The first rank accuracy rates (in percentage) of the original and fused rank lists on the Cross-Scenario (Normal versus Bag).

Angle **Segment**	**0°**	**18°**	**36°**	**54°**	**72°**	**90°**	**108°**	**126°**	**144°**	**162°**	**180°**
Body	39	23	18	15	18	18	16	15	19	26	39
Body (fused)	**70**	**61**	**55**	**38**	**46**	**38**	**44**	**37**	**52**	**61**	**65**
Head & Neck	27	25	24	28	40	44	42	38	39	34	35
Head & Neck (fused)	**66**	**71**	**73**	**75**	**85**	**89**	**84**	**76**	**78**	**72**	**71**
Shoulders	22	18	13	11	17	22	17	15	1	20	18
Shoulders (fused)	**50**	**50**	**41**	**40**	**51**	**41**	**47**	**51**	**44**	**51**	**40**
Middle	11	7	4	4	4	4	4	4	5	6	10
Middle (fused)	**31**	**22**	**13**	**8**	**10**	**5**	**12**	**7**	**10**	**14**	**25**
Lower	11	7	7	9	11	14	10	8	8	9	15
Lower (fused)	**47**	**40**	**41**	**54**	**56**	**59**	**49**	**43**	**42**	**55**	**54**
Upper Q	48	34	32	30	35	41	39	37	46	45	57
Upper Q (fused)	**90**	**80**	**83**	**79**	**68**	**72**	**75**	**79**	**83**	**85**	**88**
Upper H	40	27	23	19	21	21	18	20	27	28	39
Upper H (fused)	**66**	**58**	**53**	**42**	**37**	**37**	**35**	**43**	**55**	**56**	**63**
Torso	21	15	11	8	9	10	9	9	11	13	20
Torso (fused)	**41**	**37**	**29**	**20**	**20**	**22**	**19**	**22**	**27**	**26**	**40**
Lower H	19	11	2	7	7	8	5	6	7	11	20
Lower H (fused)	**46**	**35**	**25**	**19**	**15**	**19**	**21**	**18**	**25**	**33**	**45**

6.3.4 Rank Lists Fusion for Cross-Scenario Performance Improvements (Normal vs. Clothes)

Further rank lists fusion experiments were applied on the rank lists obtained from the cross-scenario Normal versus Clothes approach in the same manner as the experiments 4.5 . Table 6.4 shows the results.

Table 6.4: I The first rank accuracy rates (in percentage) of the original and fused rank lists on the Cross-Scenario (Normal versus Clothes).

Angle **Segment**	**0°**	**18°**	**36°**	**54°**	**72°**	**90°**	**108°**	**126°**	**144°**	**162°**	**180°**
Body	14	10	9	9	9	10	8	9	10	12	17
Body (fused)	**20**	**22**	**19**	**18**	**17**	**14**	**19**	**16**	**22**	**13**	**27**
Head & Neck	6	6	7	9	12	14	11	8	8	6	6
Head & Neck (fused)	**8**	**10**	**14**	**16**	**21**	**25**	**20**	**12**	**13**	**4**	**6**
Shoulders	3	2	3	2	2	3	3	3	3	2	3
Shoulders (fused)	**5**	**4**	**4**	**4**	**3**	**4**	**7**	**5**	**7**	**2**	**6**
Middle	4	3	3	2	2	2	2	2	3	4	6
Middle (fused)	**7**	**7**	**8**	**4**	**3**	**3**	**4**	**2**	**8**	**4**	**8**
Lower	10	7	7	7	10	10	8	7	7	8	14
Lower (fused)	**55**	**40**	**41**	**45**	**56**	**54**	**47**	**41**	**41**	**16**	**62**
Upper Q	7	8	9	10	11	10	11	12	12	9	6
Upper Q (fused)	**8**	**10**	**19**	**22**	**22**	**17**	**16**	**20**	**20**	**13**	**8**
Upper H	10	9	7	7	8	8	8	9	10	9	11
Upper H (fused)	**13**	**16**	**15**	**13**	**13**	**13**	**16**	**14**	**15**	**16**	**15**
Torso	5	4	4	3	3	4	3	3	4	4	6
Torso (fused)	**7**	**8**	**8**	**7**	**4**	**7**	**4**	**5**	**4**	**9**	**11**
Lower H	6	6	6	6	7	7	6	6	7	6	11
Lower H (fused)	**12**	**12**	**15**	**16**	**25**	**20**	**16**	**20**	**18**	**15**	**16**

The results in Table 6.4 showed that implementing the rank lists fusion on the Normal versus Clothes scenario added notable improvements to the performance of all shapes because wearing clothes, especially the winter clothes, affects large parts of the body. However, analysing the *Lower* segment results after using the rank lists fusion approach showed significant improvements compared with the improvements level of other segments, as changing the appearance by wearing different tops or coats mostly affects the upper parts of the body, whereas changing the clothes of the lower part of the body (i.e., trousers) remains within acceptable levels of shape affects.

6.4 Summary

In this Chapter, a comprehensive system output for image-based and video-based systems analyses was presented. The analysis compared the performance from both systems for each segment individually. This analysis revealed a number of findings, most importantly, that video-based system accuracy rates outperformed image-based system accuracy rates in most segments.

A developed rank lists fusion method that can be applied on the image-based generated rank lists changes that fact. This fusion improved performance for a number of different research aspects, which can be categorised into three parts: angle-based (i.e., identifying and re-identifying from the same viewing angle); inter-view scenario; and Normal versus Bag cross-scenario video sequences. A general slight performance enhancement was achieved on implementing the rank lists fusion on the Normal versus Clothes cross-scenario video sequences, and considerable improvements only on the *Lower* segment were found for this scenario.

This Chapter presented several deep analyses and fusion experiments on the results obtained from the image-based and video-based systems. These analyses and further fusion experiments open a new direction for shape-based person re-id.

Chapter 7

Conclusions and Recommendations for Future Work

This chapter includes a summary of the work completed for this thesis, followed by a discussion of the principal research findings and recommendations for future work.

7.1 Summary of the Research

The work presented in this thesis relates to intelligent surveillance and, in particular, person re-identification (re-id) applications. The shape descriptors of the proposed body segmentations were used as identifiers for person re-identification (i.e., as a unique signature for each subject). The discrimination levels of shape-based features were assessed by classifying them, using image-based and video-based approaches. The image-based system classified the signatures on a frame-by-frame basis using Linear Discriminant Analysis (LDA), which evaluated the feasibility of re-identifying subjects based on their shape static feature. The video-based approach exploited the signatures of the entire sequence (i.e., multiple frames) to re-identify subjects based on their dynamic feature occurring in the frames collection, using Dynamic Time Wrapping (DTW). The results of both systems confirmed that the shape-based features presented a high level of discrimination in person re-id application. Details of the completed work are as follows:

- Chapter 2 presented a comprehensive overview of person re-id systems and an introduction to relevant instruments. It briefly reviewed biometric systems and how person re-id fits in the overarching concept of biometric modalities and use cases. In addition, it presented a general person re-identification framework to address specific scenario-based challenges. It also included a report on several ways to represent and classify subjects and outlined deep learning methods and common evaluation metrics for the field. Finally, it provided a review of the concept of body segmentation in recognition systems, elaborating on different techniques for segmenting individuals.

- In chapter 3 the proposed framework of shape-based person re-id, including the design and implementation, was discussed. Also presented were the publicly available person re-id datasets and justification for the reason behind choosing CASIA Dataset B as the examined dataset for this research. It presented the proposed body segmentations, including the anatomical average length of each body segment based on multiple anthropological studies. In addition, it showed the arithmatec operations that algorithmically divides the human body silhouette into four suggested segments as illustrated.

- Chapter 4 presented the investigation on the discrimination level of the shape descriptors of the body and proposed body segments in the application of person re-id. The shape-based features (i.e., shape descriptors or signature) were assessed through an image-based system. This system classified these features on a frame-by-frame basis using LDA. This approach aimed to assess the feasibility of re-identifying a subject based on the subject's shape static feature. This assessment was also implemented on different data scenarios, namely, inter-view and cross-scenario. Implementation of the image-based approach in the situation in which the subject was identified and re-identified from the same angle and maintaining the same appearance outperformed a number of state-of-art systems. The results indicated that the situation in which the subject was identified and re-identified from different viewing angles (inter-view) with a change in appearance (cross-scenario) presented a comparable performance.

- Chapter 5 explored the dynamic features within the shape descriptors of the body and body segments for person re-id application. A sequence consisting of multiple frames and the shape descriptor was extracted from each frame, generating a multiple feature vectors for each sequence. This was exploited to match subjects based on their dynamic feature within the shape descriptors of the provided sequence. Therefore, this assessment approach was called a video-based approach, as the matching process was applied on a sequence-by-sequence basis using Dynamic Time Wrapping (DTW). This assessment was implemented also on different data scenarios, namely, inter-view and cross-scenario. The experimental evidence showed that implementation of the video-based approach outperformed a number of state-of-art systems. In the other scenarios, the results presented a comparable performance.

- In Chapter 6 the outcomes of the image-based system were compared with those of the video-based system, revealing significant outcomes. In addition, in this chapter a rank list fusion technique was implemented with the objective of performance enhancement. This fusion method combined the image-based system generated rank lists so that the lists generated by all frames in each sequence were replaced by one rank list. The concept of the implemented fusion method was to count the indices of each identity in the initial rank lists. The indices of each identity from each initial rank list were added together, and then the identity with the least total indices was placed in the frontal location of the new fused rank list. This method was implemented on four approaches in this research, namely, angle-based, inter-view, normal vs. bag, and normal vs. clothes scenarios. The experimental results showed a superior performance enhancement in all scenarios.

7.2 Key Findings

Based on the experimental observations completed for this thesis, the main conclusions can be summarised as follows:

- *Image-based and video-based systems' performance*. The general performance of using the body and body segments shape descriptors for person re-id was considered effective in both systems, compared to state-of-art person re-id practices. The discrimination levels (i.e., accuracy rates) generated from the video-based system outperformed the accuracy rates of the image-based system. However, as the image-based system generated multiple rank lists for each sequence (unlike the video-based system, which generated one rank list for each sequence), the fusion of one sequence rank list provided a superior performance enhancement for predicting the identity of the subject in the sequence.

- *Body segments performance*. The experiments conducted confirmed that different segments of the same body have different discrimination levels. Furthermore, some segments, such as the *Upper Half*, *Upper Quarter*, and *Head & Neck* segments, consistently outperform other segments. Other segments, such as *Lower*, *Lower Half*, and *Middle*, consistently reported low accuracy rates. Other segments presented fluctuating levels of accuracy, showing a sensitivity to a number of influencers.

- *The shape-based features influencer factors.* The experiments conducted proved that the shape descriptor is highly affected by a number of factors, most obviously body motion. However, even with the motion effects on the body, the shape descriptor still delivered discriminative information for the application of person re-id. Other important factors were the angle, which is the shooting or viewing angle from which the body was captured, and the body orientation within one sequence, where the subject changed the side of the body facing the camera. Changing the appearance by wearing different clothes, head cover, or shoes negatively affected the accuracy of the shape descriptor.

- *Rank list fusion method performance enhancement.* In addition to the superior performance enhancement that the developed fusion method introduced to the angle-based accuracy, it considerably improved the inter-view and normal vs. bag appearance scenarios as well.

7.3 Main Contributions

The main contributions of this thesis can be summarised as follows:

1. A comprehensive review on the state-of-art practices of person re-id.
2. Proposing body segmentations, with the identification of segments based on multiple anthropometry studies and using arithmetic operations.
3. Extracting the Generic Fourier Descriptor (GFD) shape descriptor from each segment.
4. Developing the image-based system to assess the static feature of the shape descriptor using LDA.
5. Developing the video-based system to assess the dynamic feature within the shape descriptor using DTW.
6. A comprehensive analysis of the systems outcomes comparing the accuracy rates for each segment from the image-based and video-based systems.
7. Proposing a rank lists fusion method seeking performance enhancements on the image-based generated rank lists.

7.4 Recommendation for Future Work

Future research can build on the work presented in this thesis to further discover the underlying aspects and improve the performance of shape-based person re-id.

First, there are wide variety of shape descriptors found in the literature, and only one shape descriptor was examined in this research. A comparable study that compares the performance of different shape descriptors will further enrich person re-id knowledge base.

Second, the review conducted in chapter 3 on the person re-id publicly available datasets showed that there is only one dataset that provides the silhouette sequence (i.e., black and white frame sequences) corresponding to the original sequence (i.e., coloured frames), which is the dataset used in this research. Also, this dataset only provides three appearance scenarios, which are normal, wearing winter clothes, and carrying a bag, whereas in real scenarios, there are additional appearance changes to consider, such as wearing a hat or head cover, different shoes, and different cultural attire. Therefore, constructing a dataset that records the subjects in additional appearance scenarios will enrich shape-based person re-id research.

Third, this research was only conducted on CASIA Dataset B for the justifications reported in Chapter 3. However, examining the same proposed systems using different datasets may make the findings of the proposed shape-based person re-id system more generalisable.

Fourth, the experimental results showed that the re-identification performance between different viewing angles (i.e., inter-view scenario) needs further enhancement. This may involve improving a metric learning that is able to match a subject's shape descriptor, even if the viewing angle is changed. In addition, what would the performance be when using the classification of the close views to classify the farther views, as the results showed that the close angles perform better than the rest viewing angles.

Fifth, the experiments conducted in this research were designed based on one of the person re-id approaches involving feature extraction and metric learning. However, the new direction of the person re-id application is to exploit the Convolutional Neural Network (CNN), as reviewed in the literature review chapter. As using the body and body segments shape descriptors is a new feature in the person re-id application, the CNN should be employed on the proposed person re-id shape-based system.

Sixth, in the literature, the extracted soft biometrics, such as clothing type and other body-related features, were fused with hard biometrics, such as face recognition. This illustrated face recognition performance enhancement introduced by each examined soft biometrics. Therefore, there is a need to investigate face recognition performance enhancement introduced by the body and body segments shape descriptors.

References

[1] X. Zhu, B. Wu, D. Huang, and W. Zheng, “Fast Open-World Person Re-Identification,” *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2286–2300, 2018, doi: 10.1109/TIP.2017.2740564.

[2] H. Ben Shitrit, J. Berclaz, F. Fleuret, and P. Fua, “Multi-Commodity Network Flow for Tracking Multiple People,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1614–1627, Aug. 2014, doi: 10.1109/TPAMI.2013.210.

[3] M. Paul, S. M. E. Haque, and S. Chakraborty, “Human detection in surveillance videos and its applications - a review,” *EURASIP J. Adv. Signal Process.*, vol. 2013, no. 1, p. 176, 2013, doi: 10.1186/1687-6180-2013-176.

[4] R. Benenson, M. Omran, J. Hosang, and B. Schiele, “Ten Years of Pedestrian Detection, What Have We Learned?,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8926, 2015, pp. 613–627.

[5] D. T. Nguyen, W. Li, and P. O. Ogunbona, “Human detection from images and videos: A survey,” *Pattern Recognit.*, vol. 51, pp. 148–175, Mar. 2016, doi: 10.1016/j.patcog.2015.08.027.

[6] J. Berclaz, F. Fleuret, E. Türetken, and P. Fua, “Multiple object tracking using k-shortest paths optimization,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 9, pp. 1806–1819, 2011, doi: 10.1109/TPAMI.2011.21.

[7] Wei Niu, Long Jiao, D. Han, and Yuan-Fang Wang, “Real-time multi-person tracking in video surveillance,” in *Fourth International Conference on*

Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint, 2003, vol. 2, pp. 1144–1148, doi: 10.1109/ICICS.2003.1292639.

[8] M. Camplani *et al.*, "Multiple human tracking in RGB-depth data: a survey," *IET Comput. Vis.*, vol. 11, no. 4, pp. 265–285, Jun. 2017, doi: 10.1049/iet-cvi.2016.0178.

[9] D. Hao Hu, S. J. Pan, V. W. Zheng, N. N. Liu, and Q. Yang, "Real world activity recognition with multiple goals," in *Proceedings of the 10th international conference on Ubiquitous computing - UbiComp '08*, 2008, p. 30, doi: 10.1145/1409635.1409640.

[10] N. Robertson and I. Reid, "A general method for human activity recognition in video," *Comput. Vis. Image Underst.*, vol. 104, no. 2–3, pp. 232–248, Nov. 2006, doi: 10.1016/j.cviu.2006.07.006.

[11] D. Gowsikhaa, S. Abirami, and R. Baskaran, "Automated human behavior analysis from surveillance videos: a survey," *Artif. Intell. Rev.*, vol. 42, no. 4, pp. 747–765, Dec. 2014, doi: 10.1007/s10462-012-9341-3.

[12] A. Bedagkar-Gala and S. K. Shah, "A survey of approaches and trends in person re-identification," *Image Vis. Comput.*, vol. 32, no. 4, pp. 270–286, Apr. 2014, doi: 10.1016/j.imavis.2014.02.001.

[13] A. K. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004, doi: 10.1109/TCSVT.2003.818349.

[14] S. Arya, N. Pratap, and K. Bhatia, "Future of Face Recognition: A Review,"

in *Procedia Computer Science*, 2015, vol. 58, pp. 578–585, doi: 10.1016/j.procs.2015.08.076.

[15] R. Jafri and H. R. Arabnia, “A Survey of Face Recognition Techniques,” *J. Inf. Process. Syst.*, vol. 5, no. 2, pp. 41–68, Jun. 2009, doi: 10.3745/JIPS.2009.5.2.041.

[16] S. Gupta and A. P. Rao, “Fingerprint Based Gender Classification Using Discrete Wavelet Transform & Artificial Neural Network,” vol. 3, no. 4, pp. 1289–1296, 2014.

[17] Z. Yao, J.-M. Le Bars, C. Charrier, and C. Rosenberger, “Literature review of fingerprint quality assessment and its evaluation,” *IET Biometrics*, vol. 5, no. 3, pp. 243–251, Sep. 2016, doi: 10.1049/iet-bmt.2015.0027.

[18] J. J. Winston and D. J. Hemanth, “A comprehensive review on iris image-based biometric system,” *Soft Comput.*, no. 1433–7479, Aug. 2018, doi: 10.1007/s00500-018-3497-y.

[19] L. Li, S. Li, S. Zhao, and L. Tan, “Research on Security of Public Security Iris Application,” in *Biometric Recognition*, 2018, pp. 459–467, doi: 10.1007/978-3-319-97909-0_49.

[20] R. Vezzani, D. Baltieri, and R. Cucchiara, “People reidentification in surveillance and forensics,” *ACM Comput. Surv.*, vol. 46, no. 2, pp. 1–37, Nov. 2013, doi: 10.1145/2543581.2543596.

[21] . W., W. Astuti, and S. Mohamed, “Intelligent Voice-Based Door Access Control System Using Adaptive-Network-based Fuzzy Inference Systems (ANFIS) for Building Security,” *J. Comput. Sci.*, vol. 3, no. 5, pp. 274–280,

2009, doi: 10.3844/jcssp.2007.274.280.

[22] X. Zhao, X. Li, Z. Wu, Y. Fu, and Y. Liu, “Multiple subcategories parts-based representation for one sample face identification,” *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 10, pp. 1654–1664, 2013, doi: 10.1109/TIFS.2013.2263498.

[23] Z. Wu, *Human Re-Identification*. Cham: Springer International Publishing, 2016.

[24] L. Zheng, S. Wang, L. Tian, Fei He, Z. Liu, and Q. Tian, “Query-adaptive late fusion for image search and person re-identification,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, vol. 07-12-June, pp. 1741–1750, doi: 10.1109/CVPR.2015.7298783.

[25] H. B. Zaman, M. H. M. Saad, M. A. Saghafi, and A. Hussain, “Review of person re-identification techniques,” *IET Comput. Vis.*, vol. 8, no. 6, pp. 455–474, Dec. 2014, doi: 10.1049/iet-cvi.2013.0180.

[26] N. Gheissari, T. B. Sebastian, P. H. Tu, J. Rittscher, and R. Hartley, “Person reidentification using spatiotemporal appearance,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, vol. 2, pp. 1528–1535, doi: 10.1109/CVPR.2006.223.

[27] R. Benenson, M. Mathias, R. Timofte, and L. Van Gool, “Pedestrian detection at 100 frames per second,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, vol. 24, pp. 2903–2910, doi: 10.1109/CVPR.2012.6248017.

[28] P. Dollar, C. Wojek, B. Schiele, and P. Perona, “Pedestrian detection: A benchmark,” in *2009 IEEE Conference on Computer Vision and Pattern*

Recognition, 2009, pp. 304–311, doi: 10.1109/CVPRW.2009.5206631.

[29] D. Gray, S. Brennan, and H. Tao, “Evaluating appearance models for recognition, reacquisition, and tracking,” *10th Int. Work. Perform. Eval. Track. Surveill. (PETS),* vol. 3, pp. 41–47, 2007.

[30] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, “Scalable Person Re-identification : A Benchmark Scalable Person Re-identification : A Benchmark,” no. December 2015, pp. 1116–1124, 2017, doi: 10.1109/ICCV.2015.133.

[31] S. S. Süsstrunk Sabine, Buckley Robert, “Standard RGB Color Spaces,” *olor Sci. Syst. Appl.*, pp. 127–134, 1999.

[32] A. K. Jain, *Fundamentals of digital image processing*. New Jersey, United States of America: Englewood Cliffs, NJ : Prentice Ha, 1989.

[33] R. Gonzalez and R. Woods, “Digital image processing and computer vision,” *Comput. Vision, Graph. Image Process.*, vol. 49, no. 1, p. 122, Jan. 1990, doi: 10.1016/0734-189X(90)90171-Q.

[34] G. H. Joblove and D. Greenberg, “Color spaces for computer graphics,” *ACM SIGGRAPH Comput. Graph.*, vol. 12, no. 3, pp. 20–25, Aug. 1978, doi: 10.1145/965139.807362.

[35] Y. Du, H. Ai, and S. Lao, “Evaluation of color spaces for person re-identification,” *Pattern Recognit. (ICPR), 2012 21st …*, no. Icpr, pp. 1371–1374, 2012.

[36] T. Z. Yuan L., “Person Re-identification Based on Color and Texture Feature Fusion,” in *ntelligent Computing Theories and Application*, 2016.

[37] N. Dalal, B. Triggs, N. Dalal, and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 886–893, 2005, doi: 10.1109/CVPR.2005.177ï.

[38] W. S. Zheng, S. Gong, and T. Xiang, “Reidentification by relative distance comparison,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 653–668, 2013, doi: 10.1109/TPAMI.2012.138.

[39] T. Avraham, I. Gurvich, M. Lindenbaum, and S. Markovitch, “Learning Implicit Transfer for Person Re-identification,” in *Computer Vision -- ECCV 2012. Workshops and Demonstrations*, 2012, pp. 381–390.

[40] L. Bazzani, M. Cristani, and V. Murino, “Symmetry-driven accumulation of local features for human characterization and re-identification,” *Comput. Vis. Image Underst.*, vol. 117, no. 2, pp. 130–144, 2013, doi: 10.1016/j.cviu.2012.10.008.

[41] M. Hirzer, P. M. Roth, K. Martin, and H. Bischof, *Relaxed Pairwise Learned Metric for Person*. Springer Berlin Heidelberg, 2012.

[42] A. Wu, W. Zheng, H. Yu, S. Gong, and J. Lai, “RGB-Infrared Cross-Modality Person Re-Identification,” *Int. Conf. Comput. Vis.*, no. ICCV, pp. 5380–5389, 2017.

[43] E. S. Jaha and M. S. Nixon, “From Clothing to Identity: Manual and Automatic Soft Biometrics,” *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 10, pp. 2377–2390, 2016, doi: 10.1109/TIFS.2016.2584001.

[44] E. S. Jaha and M. S. Nixon, “Analysing Soft Clothing Biometrics for Retrieval,” in *Biometric Authentication. BIOMET 2014. Lecture Notes in*

Computer Science, Springer, Cham, 2014, pp. 234–245.

[45] E. S. Jaha and M. S. Nixon, "Soft biometrics for subject identification using clothing attributes," in *IEEE International Joint Conference on Biometrics*, 2014, pp. 1–6, doi: 10.1109/BTAS.2014.6996278.

[46] J. Shutler, M. Grant, M. S. Nixon, and J. N. Carter, "On a Large Sequence-Based Human Gait Database," *Proc. Fourth Int. Conf. Recent Adv. Soft Comput.*, pp. 66–72, 2002, doi: 10.1007/978-3-540-45240-9_46.

[47] T. Joachims, "Optimizing search engines using clickthrough data," *Proc. eighth ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '02*, p. 133, 2002, doi: 10.1145/775066.775067.

[48] S. Samangooei and M. S. Nixon, "Performing content-based retrieval of humans using gait biometrics," *Multimed. Tools Appl.*, vol. 49, no. 1, pp. 195–212, 2010, doi: 10.1007/s11042-009-0391-8.

[49] M. S. Nixon, "Viewpoint Invariant Subject Retrieval via Soft Clothing Biometrics," *2015 Int. Conf. Biometrics*, pp. 73–78, 2015.

[50] M. S. Nixon, B. H. Guo, S. V Stevenage, E. S. Jaha, N. Almudhahka, and D. Martinho-Corbishley, "Towards automated eyewitness descriptions: describing the face, body and clothing for recognition," *Vis. cogn.*, vol. 25, no. 4–6, pp. 524–538, 2017, doi: 10.1080/13506285.2016.1266426.

[51] S. Ojha and S. Sakhare, "Image processing techniques for object tracking in video surveillance- A survey," *2015 Int. Conf. Pervasive Comput.*, vol. 00, no. c, pp. 1–6, 2015, doi: 10.1109/PERVASIVE.2015.7087180.

[52] D. Zhang and G. Lu, "Review of shape representation and description

techniques," *Pattern Recognit.*, vol. 37, no. 1, pp. 1–19, 2004, doi: 10.1016/j.patcog.2003.07.008.

[53] E. Gonzalez-Sosa, R. Vera-Rodriguez, J. Fierrez, and V. M. Patel, "Person Recognition beyond the Visible Spectrum: Combining Body Shape and Texture from mmW Images," *2018 Int. Conf. Biometrics*, pp. 241–246, 2018, doi: 10.1109/ICB2018.2018.00044.

[54] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002, doi: 10.1109/34.993558.

[55] E. Poongothai and A. Suruliandi, "COLOUR, TEXTURE AND SHAPE FEATURE ANALYSIS FOR PERSON RE-IDENTIFICATION TECHNIQUE," *Indian J. Sci. Technol.*, vol. 9, no. 29, pp. 17–26, Aug. 2016, doi: 10.17485/ijst/2016/v9i29/93823.

[56] F. Mokhtarian and A. Mackworth, "Scale-Based Description and Recognition of Planar Curves and Two-Dimensional Shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 1, pp. 34–43, 1986, doi: 10.1109/TPAMI.1986.4767750.

[57] E. Gonzalez-Sosa, R. Vera-Rodriguez, J. Fierrez, and V. M. Patel, "Exploring Body Shape From mmW Images for Person Recognition," *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 9, pp. 2078–2089, Sep. 2017, doi: 10.1109/TIFS.2017.2695979.

[58] O. A. Arigbabu, S. M. S. Ahmad, W. A. W. Adnan, S. Yussof, V. Iranmanesh, and F. L. Malallah, "Estimating Body Related Soft Biometric Traits in Video Frames," *Sci. World J.*, vol. 2014, pp. 1–13, 2014, doi: 10.1155/2014/460973.

[59] A. Criminisi, I. Reid, and A. Zisserman, “Single View Metrology,” *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 123–148, Nov. 2000, doi: 10.1023/A:1026598000963.

[60] A. Dantcheva, C. Velardo, A. D’Angelo, and J. L. Dugelay, “Bag of soft biometrics for person identification: New trends and challenges,” *Multimed. Tools Appl.*, vol. 51, no. 2, pp. 739–777, 2011, doi: 10.1007/s11042-010-0635-7.

[61] C. BenAbdelkader, R. Cutler, and L. Davis, “Person identification using automatic height and stride estimation,” *Object Recognit. Support. by user Interact. Serv. Robot.*, vol. 4, pp. 377–380, 2002, doi: 10.1109/ICPR.2002.1047474.

[62] N. H. Nguyen and R. Hartley, “Height measurement for humans in motion using a camera: A comparison of different methods,” *2012 Int. Conf. Digit. Image Comput. Tech. Appl. DICTA 2012*, pp. 1–8, 2012, doi: 10.1109/DICTA.2012.6411679.

[63] E. Jeges, I. Kispál, and Z. Hornák, “Measuring human height using calibrated cameras,” *2008 Conf. Hum. Syst. Interact. HSI 2008*, pp. 755–760, 2008, doi: 10.1109/HSI.2008.4581536.

[64] D. M. Hansen, B. K. Mortensen, P. T. Duizer, J. R. Andersen, and T. B. Moeslund, “Automatic Annotation of Humans in Surveillance Video,” in *Fourth Canadian Conference on Computer and Robot Vision (CRV ’07)*, 2007, pp. 473–480, doi: 10.1109/CRV.2007.12.

[65] O. A. Arigbabu, S. M. S. Ahmad, W. A. W. Adnan, and S. Yussof, “Integration of Multiple Soft Biometrics for Human Identification,” *Pattern Recognit. Lett.*,

vol. 68, pp. 278–287, 2015, doi: 10.1016/j.patrec.2015.07.014.

[66] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, “Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition,” in *CVPR 2011 WORKSHOPS*, 2011, pp. 74–81, doi: 10.1109/CVPRW.2011.5981881.

[67] F. Lv, T. Zhao, and R. Nevatia, “Camera calibration from video of a walking human,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1513–1518, 2006, doi: 10.1109/TPAMI.2006.178.

[68] K. Z. Lee, “A simple calibration approach to single view height estimation,” *Proc. 2012 9th Conf. Comput. Robot Vision, CRV 2012*, pp. 161–166, 2012, doi: 10.1109/CRV.2012.29.

[69] P. Tome, J. Fierrez, R. Vera-Rodriguez, and M. S. Nixon, “Soft Biometrics and Their Application in Person Recognition at a Distance,” *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 3, pp. 464–475, Mar. 2014, doi: 10.1109/TIFS.2014.2299975.

[70] R. Vera-Rodriguez, P. Marin-Belinchon, E. Gonzalez-Sosa, P. Tome, and J. Ortega-Garcia, “Exploring automatic extraction of body-based soft biometrics,” *Proc. - Int. Carnahan Conf. Secur. Technol.*, vol. 2017-Octob, pp. 1–6, 2017, doi: 10.1109/CCST.2017.8167841.

[71] A. Lotfi and J. M. Garibaldi, *Applications and Science in Soft Comuting, Series*, no. January 2004. Springer, Berlin, Heidelberg, 2004.

[72] M. S. Nixon, P. L. Correia, K. Nasrollahi, T. B. Moeslund, A. Hadid, and M. Tistarelli, “On soft biometrics,” *Pattern Recognit. Lett.*, vol. 68, pp. 218–230,

Dec. 2015, doi: 10.1016/j.patrec.2015.08.006.

[73] A. Dantcheva, P. Elia, and A. Ross, “What Else Does Your Biometric Data Reveal? A Survey on Soft Biometrics,” *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 3, pp. 441–467, Mar. 2016, doi: 10.1109/TIFS.2015.2480381.

[74] Z. Liu, Z. Zhang, Q. Wu, and Y. Wang, “Enhancing Person Re-identification by Integrating Gait Biometric,” in *Computer Vision - ACCV 2014 Workshops*, 2015, pp. 35–45.

[75] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, “Person Re-Identification by Symmetry Driven Accumulation of Local Features,” *Proc. Comput. Vis. Pattern Recognit.*, pp. 2360–2367, 2010.

[76] K. Weinberger, “Distance metric learning for large margin nearest neighbor classification,” *J. Mach. Learn. Res.*, pp. 207–244, 2005, doi: 10.1142/S021800141100897X.

[77] X. Ma *et al.*, “Person re-identification by unsupervised video matching,” *Pattern Recognit.*, vol. 65, pp. 197–210, May 2017, doi: 10.1016/j.patcog.2016.11.018.

[78] O. Javed, K. Shafique, Z. Rasheed, and M. Shah, “Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views,” *Comput. Vis. Image Underst.*, vol. 109, no. 2, pp. 146–162, 2008, doi: 10.1016/j.cviu.2007.01.003.

[79] D. Makris, T. Ellis, and J. Black, “Bridging the gaps between cameras,” *Proc. 2004 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, 2004. CVPR 2004.*, vol. 2, pp. 205–210, 2004, doi: 10.1109/CVPR.2004.1315165.

[80] R. Mazzon, S. F. Tahir, and A. Cavallaro, “Person re-identification in crowd,” *Pattern Recognit. Lett.*, vol. 33, no. 14, pp. 1828–1837, 2012, doi: 10.1016/j.patrec.2012.02.014.

[81] I. O. De Oliveira and J. L. D. S. Pio, “People reidentification in a camera network,” *8th IEEE Int. Symp. Dependable, Auton. Secur. Comput. DASC 2009*, pp. 461–466, 2009, doi: 10.1109/DASC.2009.33.

[82] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux, “Interest Points Harvesting in Video Sequences for Efficient Person Identification,” *Proc. 10th Eur. Conf. Comput. Vis.*, 2008.

[83] O. Hamdoun *et al.*, “PERSON RE-IDENTIFICATION IN MULTI-CAMERA SYSTEM BY SIGNATURE BASED ON INTEREST POINT DESCRIPTORS COLLECTED ON SHORT VIDEO SEQUENCES Omar Hamdoun , Fabien Moutarde , Bogdan Stanciulescu and Bruno Steux Mines ParisTech 60 Bd St Michel , F-75006 Paris , FRAN,” *Robotics*, pp. 0–5, 2008.

[84] A. Albiol, A. Albiol, J. M. Mossi, and J. Oliver, “Who is who at different cameras: people re-identification using depth cameras,” *IET Comput. Vis.*, vol. 6, no. 5, pp. 378–387, 2012, doi: 10.1049/iet-cvi.2011.0140.

[85] A. D’Angelo and J.-L. Dugelay, “People re-identification in camera networks based on probabilistic color histograms,” 2011, vol. 33, no. 0, p. 78820K, doi: 10.1117/12.876453.

[86] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, “Large scale metric learning from equivalence constraints,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, no. Ldml, pp. 2288–2295, 2012, doi: 10.1109/CVPR.2012.6247939.

[87] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by Local Maximal Occurrence representation and metric learning," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 2197–2206, 2015, doi: 10.1109/CVPR.2015.7298832.

[88] W. S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 653–668, 2013, doi: 10.1109/TPAMI.2012.138.

[89] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised Salience Learning for Person Re-identification," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3586–3593, doi: 10.1109/CVPR.2013.460.

[90] X. Liu, H. Wang, Y. Wu, J. Yang, and M. H. Yang, "An ensemble color model for human re-identification," *Proc. - 2015 IEEE Winter Conf. Appl. Comput. Vision, WACV 2015*, pp. 868–875, 2015, doi: 10.1109/WACV.2015.120.

[91] D. Gray and H. Tao, "Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features," in *Computer Vision -- ECCV 2008*, 2008, pp. 262–275.

[92] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep Filter Pairing Neural Network for Person Re-identification," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 152–159, doi: 10.1109/CVPR.2014.27.

[93] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep Metric Learning for Person Re-identification," in *2014 22nd International Conference on Pattern Recognition*, 2014, vol. 83, no. 12, pp. 34–39, doi: 10.1109/ICPR.2014.16.

[94] L. Zheng, Y. Yang, and A. G. Hauptmann, “Person Re-identification: Past, Present and Future,” *Acad. Pediatr.*, vol. 11, no. 3, pp. 234–239, Oct. 2016, doi: 10.1016/j.acap.2010.12.001.

[95] H. Jansen, M. P. Gallee, and F. H. Schroder, “Analysis of sonographic pattern in prostatic cancer: Comparison of longitudinal and transversal transrectal ultrasound with subsequent radical prostatectomy specimens,” *Eur. Urol.*, vol. 18, no. 3, pp. 174–178, 1990, doi: 10.1159/000463903.

[96] E. Ahmed, M. Jones, and T. K. Marks, “An improved deep learning architecture for person re-identification,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3908–3916, doi: 10.1109/CVPR.2015.7299016.

[97] R. R. Varior, B. Shuai, J. Lu, D. Xu, and G. Wang, “A Siamese Long Short-Term Memory Architecture for Human Re-Identification,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9911 LNCS, pp. 135–153, Jul. 2016, doi: 10.1007/978-3-319-46478-7_9.

[98] A. Khan, J. Zhang, and Y. Wang, “Appearance-based re-identification of people in video,” *Proc. - 2010 Digit. Image Comput. Tech. Appl. DICTA 2010*, pp. 357–362, 2010, doi: 10.1109/DICTA.2010.67.

[99] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, “Person re-identification by symmetry-driven accumulation of local features,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2360–2367, 2010, doi: 10.1109/CVPR.2010.5539926.

[100] S. Bąk, E. Corvee, F. Brémond, and M. Thonnat, “Person re-identification

using spatial covariance regions of human body parts," *Proc. - IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2010*, pp. 435–440, 2010, doi: 10.1109/AVSS.2010.34.

[101] A. Bedagkar-Gala and S. K. Shah, "Part-based spatio-temporal model for multi-person re-identification," *Pattern Recognit. Lett.*, vol. 33, no. 14, pp. 1908–1915, 2012, doi: 10.1016/j.patrec.2011.09.005.

[102] D. Zhang and G. Lu, "Shape-based image retrieval using generic Fourier descriptor," *Signal Process. Image Commun.*, vol. 17, no. 10, pp. 825–848, Nov. 2002, doi: 10.1016/S0923-5965(02)00084-X.

[103] Z. Xie, L. Li, X. Zhong, and L. Zhong, "Image-to-Video Person Re-Identification by Reusing Cross-modal Embeddings," *Pattern Recognit. Lett.*, pp. 1–7, 2018.

[104] M. Welling, "Fisher Linear Discriminant Analysis," *Science (80-.).*, vol. 1, no. 2, pp. 1–3, 2009, doi: 10.1109/NNSP.1999.788121.

[105] T. K. Vintsyuk, "Speech discrimination by dynamic programming," *Cybernetics*, vol. 4, no. 1, pp. 52–57, Jan. 1968, doi: 10.1007/BF01074755.

[106] M. Bellare and P. Rogaway, "The Exact Security of Digital Signatures-How to Sign with RSA and Rabin," in *Lecture Notes in Computer Science*, 1996, pp. 399–416.

[107] W.-S. Zheng, S. Gong, and T. Xiang, "Associating Groups of People," in *Procedings of the British Machine Vision Conference 2009*, 2009, vol. 5, no. 1, pp. 23.1-23.11, doi: 10.5244/C.23.23.

[108] L. B. and V. M. Dong Seon Cheng, Marco Cristani, Michele Stoppa, "Custom

Pictorial Structures for Re-identification," *Bmvc*, pp. 68.1--68.11, 2011, doi: http://dx.doi.org/10.5244/C.25.68.

[109] D. Tan, K. Huang, S. Yu, and T. Tan, "Efficient night gait recognition based on template matching," *Proc. - Int. Conf. Pattern Recognit.*, vol. 3, pp. 1000–1003, 2006, doi: 10.1109/ICPR.2006.478.

[110] I. I. Conference and I. Processing, "EVALUATION FRAMEWORK ON TRANSLATION-INVARIANT REPRESENTATION FOR CUMULATIVE FOOT PRESSURE IMAGE Shuai Zheng , Kaiqi Huang , Tieniu Tan National Laboratory of Pattern Recognition , Institute of Automation , Chinese Academy of Sciences," pp. 205–208, 2011.

[111] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7724 LNCS, no. PART 1, pp. 31–44, 2013, doi: 10.1007/978-3-642-37331-2_3.

[112] S. Wang, M. Lewandowski, J. Annesley, and J. Orwell, "Re-identification of pedestrians with variable occlusion and scale," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 1876–1882, 2011, doi: 10.1109/ICCVW.2011.6130477.

[113] L. Ma, H. Liu, L. Hu, C. Wang, and Q. Sun, "Orientation Driven Bag of Appearances for Person Re-identification," pp. 1–13, 2016.

[114] C. C. Loy, C. Liu, and S. Gong, "Person re-identification by manifold ranking," *2013 IEEE Int. Conf. Image Process. ICIP 2013 - Proc.*, no. i, pp. 3567–3571, 2013, doi: 10.1109/ICIP.2013.6738736.

[115] D. Baltieri, R. Vezzani, and R. Cucchiara, "3DPeS," in *Proceedings of the*

2011 joint ACM workshop on Human gesture and behavior understanding - J-HGBU '11, 2011, p. 59, doi: 10.1145/2072572.2072590.

[116] W. Li and X. Wang, "Locally Aligned Feature Transforms across Views," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3594–3601, doi: 10.1109/CVPR.2013.461.

[117] A. Das, A. Chakraborty, and A. K. Roy-Chowdhury, "Consistent Re-identification in a Camera Network," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8690 LNCS, no. PART 2, 2014, pp. 330–345.

[118] L. Zheng, H. Zhang, S. Sun, M. Chandraker, Y. Yang, and Q. Tian, "Person Re-identification in the Wild," 2016, doi: 10.1109/CVPR.2017.357.

[119] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, "End-to-End Deep Learning for Person Search," *arXiv cs.CV*, vol. 4, p. 01850, 2016, doi: 10.1109/CVPR.2017.360.

[120] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," pp. 3754–3762.

[121] S. Karanam, M. Gou, Z. Wu, A. Rates-Borras, O. Camps, and R. J. Radke, "A Systematic Evaluation and Benchmark for Person Re-Identification: Features, Metrics, and Datasets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[122] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person Transfer GAN to Bridge Domain Gap for Person Re-Identification," 2017, doi: 10.1109/CVPR.2018.00016.

[123] Shiqi Yu, Daoliang Tan, and Tieniu Tan, “A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition,” in *18th International Conference on Pattern Recognition (ICPR’06)*, 2006, vol. 4, pp. 441–444, doi: 10.1109/ICPR.2006.67.

[124] W. R. Schwartz and L. S. Davis, “Learning Discriminative Appearance-Based Models Using Partial Least Squares,” in *2009 XXII Brazilian Symposium on Computer Graphics and Image Processing*, 2009, pp. 322–329, doi: 10.1109/SIBGRAPI.2009.42.

[125] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, “Person re-identification by descriptive and discriminative classification,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6688 LNCS, pp. 91–102, 2011, doi: 10.1007/978-3-642-21227-7_9.

[126] N. Martinel and C. Micheloni, “Re-identify people in wide area camera network,” in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2012, pp. 31–36, doi: 10.1109/CVPRW.2012.6239203.

[127] A. Bialkowski, S. Denman, S. Sridharan, C. Fookes, and P. Lucey, “A database for person re-identification in multi-camera surveillance networks,” *2012 Int. Conf. Digit. Image Comput. Tech. Appl. DICTA 2012*, 2012, doi: 10.1109/DICTA.2012.6411689.

[128] T. Wang, S. Gong, X. Zhu, and S. Wang, “Person Re-Identification by Discriminative Selection in Video Ranking,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 12, pp. 2501–2514, 2016, doi: 10.1109/TPAMI.2016.2522418.

[129] D. F. B, M. Taiana, A. Nambiar, J. Nascimento, and A. Bernardino, “The HDA + Data Set for Research on Fully Automated Re-identification Systems,” pp. 241–255, 2015, doi: 10.1007/978-3-319-16199-0.

[130] Y. Kawanishi, Y. Wu, M. Mukunoki, and M. Minoh, “Shinpuhkan2014: A Multi-Camera Pedestrian Dataset for Tracking People across Multiple Cameras,” *20th Korea-Japan Jt. Work. Front. Comput. Vision, FCV2014*, pp. 1–5, 2014.

[131] L. Zheng *et al.*, “MARS: A Video Benchmark for Large-Scale Person Re-Identification,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9910 LNCS, 2016, pp. 868–884.

[132] G. A. Jackson, “Survey of EMC measurement techniques,” *Electron. Commun. Eng. J.*, vol. 1, no. 2, p. 61, 1989, doi: 10.1049/ecej:19890011.

[133] W. T. Dempster and G. R. L. CGaughran, “Properties of body segnments base on size and weight,” *Am. J. Anat.*, vol. 120, no. 7414, pp. 33–54, 1889.

[134] R. Contini, R. J. Drillis, and M. Bluestein, “Determination of Body Segment Parameters,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 5, no. 5, pp. 493–504, Oct. 1963, doi: 10.1177/001872086300500508.

[135] D. L. P., “ADJUSTMENTS TO ZATSIORSKY-SELUYANOV’S SEGMENT IN ERTIA PARAMETERS.” pp. 1223–1230, 1996, doi: 0021-9290(95)00178-6.

[136] I. Venkat and P. De Wilde, “Robust gait recognition by learning and exploiting sub-gait characteristics,” *Int. J. Comput. Vis.*, vol. 91, no. 1, pp. 7–23, 2011,

doi: 10.1007/s11263-010-0362-6.

[137] M. S. B. Hu, H. Jin, J. Wang, and E. Keogh, “Generalizing DTW to the multi-dimensional case requires an adaptive approach,” *Data Min. Knowl. Discov.*, vol. 31, no. 1, pp. 1–31, 2017, doi: 10.1007/s10618-016-0455-0.

[138] S. Wu, Y.-C. Chen, X. Li, A.-C. Wu, J.-J. You, and W.-S. Zheng, “An Enhanced Deep Feature Representation for Person Re-identification,” *2016 IEEE Winter Conf. Appl. Comput. Vis.*, pp. 1–8, Apr. 2016, doi: 10.1109/WACV.2016.7477681.

[139] Y. Yang, X. Liu, Q. Ye, and D. Tao, “Ensemble Learning-Based Person Re-identification with Multiple Feature Representations,” *Complexity*, vol. 2018, pp. 1–12, Sep. 2018, doi: 10.1155/2018/5940181.

[140] T. kam Ho, J. J. Hull, and S. N. Srihari, “Decision Combination in Multiple Classifier Systems,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 1, pp. 66–75, 1994, doi: 10.1109/34.273716.

[141] A. Kumar and S. Shekhar, “Personal identification using multibiometrics rank-level fusion,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 41, no. 5, pp. 743–752, 2011, doi: 10.1109/TSMCC.2010.2089516.

[142] Q. Leng, R. Hu, C. Liang, Y. Wang, and J. Chen, “Person re-identification with content and context re-ranking,” *Multimed. Tools Appl.*, vol. 74, no. 17, pp. 6989–7014, 2015, doi: 10.1007/s11042-014-1949-7.