Investigating Adaptive Multi-modal Approaches for Person Identity Verification Based on Face and Gait Fusion

S.M. Emdad Hossain

A Thesis Submitted in Partial Fulfilment of the Requirements of the Degree of Doctor of Philosophy
Faculty of Education Science Technology and Mathematics

September 2014
Abstract
With my thesis, I have established a novel human-identification scheme from long range face-gait profiles in surveillance videos. I investigated the role of multi view face-gait images acquired from multiple cameras, the importance of surveillance and visible range images in ascertaining identity, the impact of multimodal fusion, and efficient subspace features and classifier methods, and along with side face-ear biometric traits; the role of soft/secondary biometric (walking style) in enhancing the accuracy and robustness of the identification systems. An extensive, experimental evaluation of several subspace based side face-ear, gait feature extraction approaches and learning classifier methods on different datasets from publicly available databases (CASIA-China, Human Action Database- Sweden and UCMG Database-University of Canberra) has shown a significant improvement in recognition accuracy and robustness with multimodal fusion of multi-view face-ear, gait images from visible and infrared cameras acquired from different video surveillance scenarios.
Acknowledgement
I would like to express my sincere gratitude to my supervisor A/Prof. Dr Girija Chetty for the continuous support to my PhD study and research, for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor for my PhD study. Further, I also would like to thank the rest of my supervisory panel, Professor Michael Wagner and A/Prof. Roland Goecke for their encouragement, and insightful comments. My sincere thanks also goes to all faculty members, especially the Dean Professor Geoffrey Riordan, Associate Dean (Research) Professor Dharmendra Sharma, Professor Ali Quazi, Serena Chong, Kylie Reece, Coral Suthern, Jason Weber, and all members of previous faculty ISE (Information Science and Engineering) for their endless support to my research-project.

I also like to thank my fellow research companions, the Research Student office, graduate research office, UC-marketing office, specially Claudia Donnan and Michelle Mcaulay, Prof. Robert Fitzgerald from INSPIRE centre, UC-Library and my friends who were giving me continuous support and inspiration on this research.

Last but not the least; I would like to thank my family: my dear father Late Shafiul Alam and dear mother Firoza Begum, all of my brothers and only sister for supporting me mystically throughout my life.
Table of Contents

Abstract ................................................................................................................................. iii
Acknowledgement .................................................................................................................. vii
List of Table .......................................................................................................................... 13
List of Figure ......................................................................................................................... 15

Chapter 1: Introduction ........................................................................................................... 1

Chapter 2: Literature Review .................................................................................................. 9
2.1 Introduction ...................................................................................................................... 9
2.2 Physiological Approach ................................................................................................. 9
2.3 Behavioural Approach .................................................................................................... 10
2.4 Computational Approach .............................................................................................. 11
2.5 Combined Approach ...................................................................................................... 17
2.6 Summary ......................................................................................................................... 18

Chapter 3: Face-Gait Fusion with Human Action Recognition Dataset ................................... 19
3.1 Introduction ..................................................................................................................... 19
3.2 Background .................................................................................................................... 20
3.3 Methodology .................................................................................................................. 24
3.4 Motivation for Face-Gait Fusion .................................................................................... 27
3.5 Principle Component Analysis (PCA) ............................................................................. 28
3.6 Linear Discriminant Analysis (LDA) ............................................................................. 30
3.7 Experiments .................................................................................................................... 32
  3.7.1 Feature Level Fusion ............................................................................................... 32
  3.7.2 Score Level Fusion ................................................................................................. 36
  3.7.3 Holistic vs. Hierarchical Fusion ............................................................................. 39
3.8 Fusion of Ear, Side-face and Gait .................................................................................. 43
  3.8.1 Experiments and Results: .................................................................................... 43
3.9 Summary ......................................................................................................................... 50

Chapter 4: Multimodal Face-Gait Fusion with CASIA Gait Database .................................... 53
4.1 Introduction ..................................................................................................................... 53
4.2 Literature Search .......................................................................................................... 54
4.3 Methodology .................................................................................................................. 56
  4.3.1 Multi-Layer Perceptron ....................................................................................... 57
  4.3.2 J48 Classifier ....................................................................................................... 58
Chapter 5: Development of University of Canberra Multimodal Gait (UCMG) Database

5.1 Introduction .............................................................................................................. 101
5.2 Literature Search for UCMG-Database ................................................................. 102
5.3 MMB Dataset ........................................................................................................... 105
  5.3.1 Normal Walking ................................................................................................. 106
  5.3.2 Fast Walking ....................................................................................................... 106
  5.3.3 Normal Walking with Heavy Bag ........................................................................ 107
  5.3.4 Normal Walking with Long Jacket (Overcoat) .................................................... 108
  5.3.5 Normal Walking with Hat .................................................................................. 108
  5.3.6 Normal Walking with Hoody ............................................................................ 109
  5.3.7 Normal Walking with Mask .............................................................................. 109
5.4 Summary .................................................................................................................... 110
Chapter 6: Multimodal Face-Gait-Ear Fusion with UCMG Database .................................................. 111

6.1 Introduction ................................................................................................................................. 111

6.2 Background ................................................................................................................................. 112

6.3 Methodology .............................................................................................................................. 114

   6.3.1 Principle Component Analysis .............................................................................................. 114

   6.3.2 Linear Discriminant Analysis ............................................................................................... 115

   6.3.3 Multilayer Perceptron (MLP) ............................................................................................. 115

6.4 Experiment ................................................................................................................................. 116

6.5 Experiment and comparison with CASIA Gait Database and UCM Gait Database .................. 121

   6.5.1 Methodology ....................................................................................................................... 122

   6.5.1.1 Variant of PCA Feature (v-PCA) ..................................................................................... 122

   6.5.1.2 Deep Learning Feature (DLF) ......................................................................................... 124

6.6 Experiment ................................................................................................................................. 125

6.7 Summary .................................................................................................................................... 127

Chapter 7: Conclusions and Further Work ..................................................................................... 129

Bibliography ..................................................................................................................................... 131
List of Table

Table 1: PCA with Bayesian Classifiers and 1-Nearest Neighbour Classifier .................................................. 34
Table 2: PCA with face-gait fusion with Bayesian Classifiers and 1-Nearest Neighbour Classifier ........... 35
Table 3: LDA with Bayesian Classifiers and 1-Nearest Neighbour Classifier ........................................... 35
Table 4: LDA face - gait fusion with Bayesian Classifiers and 1-Nearest Neighbour Classifier .............. 36
Table 5: PCA with Bayesian Classifiers and 1-Nearest Neighbour Classifier ........................................... 36
Table 6: PCA-LDA with Bayesian Classifiers and 1-Nearest Neighbour Classifier ................................ 39
Table 7: Holistic vs hierarchical Fusion ........................................................................................................ 41
Table 8: Ear recognition rate by using PCA-LDA .......................................................................................... 46
Table 9: Fusion result of ear and face ............................................................................................................. 46
Table 10: Gait-only recognition rate by using PCA-LDA .............................................................................. 47
Table 11: Result of feature level fusion of ear, face and gait ....................................................................... 48
Table 12: Influence of dimensionality of PCA-LDA features on the accuracy ............................................. 65
Table 13: Identification in LDA-Bagging protocol ......................................................................................... 81
Table 14: Classifier performance for visible range dataset (Dataset B) with PCA feature with 50 dimensions (NB – naive Bayes; MLP – Multilayer Perceptron; TRP – True Positive Rate and FPR – False Positive Rate) ................................................................................................................................. 88
Table 15: Classifier performance for visible range dataset (Dataset B) with LDA feature with 50 dimensions (SVM-L- Support Vector Machine-Linear Kernel; RBF- Radial Basis Function Kernel; Poly – Polynomial Kernel; SMO – Sequential Minimal Optimization) ................................................................................................................................. 89
Table 16: Classifier performance for Infrared range dataset (Dataset C) with LDA feature with 50 dimensions (SVM-L- Support Vector Machine-Linear Kernel; RBF- Radial Basis Function Kernel; Poly – Polynomial Kernel; SMO – Sequential Minimal Optimization) ................................................................................................................................. 90
Table 17: Classifier performance for fusion of visible and infrared image ................................................. 91
Table 18: Result of fusion of normal walking and fast walking ................................................................. 92
Table 19: Recognition Accuracy for Multi-View Fusion (3D) ................................................................. 96
Table 20: Recognition accuracy for cross-camera feature fusion ............................................................ 97
Table 21: Recognition accuracy with different walking pattern ............................................................ 98
Table 22: List of popular database and their modality ............................................................................ 103
Table 23: Summary of UCMG-Database ..................................................................................................... 105
Table 24: Identification with LDA-MLP ....................................................................................................... 117
Table 25: Performance of Learning Features (dimension = 20) with MLP classifier ............................ 125
Table 26: Performance of Learning Features (dimension = 20) for different classifiers ....................... 126
Table 27: Performance of Decision Fusion of LDF and DLF Features (dimension = 20) ....................... 127
List of Figure

Figure 1: Sample images from human action database for walking sequences [36] ......................................................... 25
Figure 2: Sample images from multiple frames of a single person’s walking sequence [36] ........................................ 25
Figure 3: Schematic for proposed multimodal identification scheme based on fusion of side face and gait cues extracted from low-resolution video .......................................................... 26
Figure 4: Eigen face of side-face only .......................................................................................................................... 33
Figure 5: Eigen face of gait-only (lower part of the body) [23] ................................................................................. 34
Figure 6: Eigen face of full gait ................................................................................................................................. 34
Figure 7: Eigen-Faces of side-face ............................................................................................................................ 38
Figure 8: Eigen face of only-gait [50] ....................................................................................................................... 38
Figure 9: Holistic fusion (Score-level fusion with equal weightage for face and gait scores for proposed multimodal identification scheme) .................................................................................. 40
Figure 10: Hierarchical Fusion scheme with Gait classifier as 1st stage and Face Classifier as a 2nd stage classifier .......................................................................................................................... 41
Figure 11: Principal Eigen values for first two principal components with PCA –LDA and gait only data set ................................................................. 42
Figure 12: Scatter plot for first two principal components with PCA –LDA and gait only data set ..... 42
Figure 13: Sample image extracted from video sequence ..................................................................................... 44
Figure 14: Process flow of the experiment .................................................................................................................. 44
Figure 15: PCA Eigen Value of ear recognition ........................................................................................................ 46
Figure 16: Scatter graph of Eigen value of face-ear fusion .................................................................................... 47
Figure 17: Eigen face of projected gait ....................................................................................................................... 48
Figure 18: CCA (Canonical correlation analysis) Map of Eigen value ..................................................................... 49
Figure 19: k-mean clustering of projected Eigen value .......................................................................................... 49
Figure 20: Sample images ........................................................................................................................................ 55
Figure 21: Extracted silhouettes .............................................................................................................................. 55
Figure 22: The extracted feature (Eigen value) using PCA in Dataset B .............................................................. 57
Figure 23: Perceptron network with three layers [58] .............................................................................................. 58
Figure 24: Normalized feature vectors ................................................................................................................... 60
Figure 25: Accuracy difference in applied classifiers ............................................................................................. 61
Figure 26: Accuracy differences in PCA-LDA with MLP ....................................................................................... 61
Figure 27: Accuracy difference in PCA-LDA with MLP ....................................................................................... 62
Figure 28: Accuracy difference in different folds LDA-MLP with dataset B ........................................................... 63
Figure 29: Accuracy difference in different folds LDA-MLP with Dataset C ......................................................... 64
Figure 30: Extracted silhouettes in 126 degree and 90 degree ............................................................................... 66
Figure 31: Extracted Eigen value ............................................................................................................................ 67
Figure 32: Combined Logistic Function .................................................................................................................. 68
Figure 33: Identification in 90 degree view point ..................................................................................................... 69
Figure 34: Identification in 126 degree view point .................................................................................................. 69
Figure 35: Person Identification in Fusion .............................................................................................................. 70
Figure 36: LDA values extracted from silhouettes ................................................................................................. 72
Figure 37: MLP network architecture for the proposed scheme ......................................................................... 73
Figure 38: Rate of Identification in 36 degree view point ........................................................................................ 74
Figure 39: Rate of Identification in 90 degree view point ....................................................................................... 75
Figure 40: Rate of Identification in 126 degree view point .................................................................................... 75
Chapter 1: Introduction

Automatic identity recognition of persons based on video cues is a highly challenging pattern recognition and machine learning technology that allows machines to automatically recognise and identify humans based on learning patterns in data corresponding to their physiological or behavioural biometric information. It is distinct from, although related to, statistics, and it can be differentiated by its focus on creating technology rather than the human-centred analysis of data.

Video surveillance in public places and facilities has become omnipresent, and has become the first line of defence for protecting assets and people for different types of operating scenarios and applications – be it a civilian public space for access control to a facility, or financial and transaction oriented applications, or the high security immigration and border control check points. It has become an enabler of trust, integrity and security in the new Digital Economy. In the wake of continuing terror-events worldwide, the surveillance technologies where user co-operation needed is minimal (video-based) are assuming more important role in safeguarding the security of the people, assets, and borders. Vast quantities of surveillance data are now routinely collected and stored because it is affordable to do so. The building blocks of such automated surveillance systems are based on models and algorithms that are implemented as several computational processing stages in sequence, such as pre-processing, feature extraction, feature selection, learning and recognition that together try to make sense of the huge surveillance data flood [1].

Current state-of-the-art video surveillance systems, when used for recognizing the identity of the person in the scene based on face images, cannot perform very well. This is because face recognition models and algorithms proposed so far were developed and tested for constrained environments. These models and algorithms are not designed to cope with CCTV
surveillance data in which face images are captured unconstrained from a distance. Also, though much progress has been made in the past decade on visual based automatic person identification through utilizing different biometrics, including face, voice and fingerprints, each of these modalities work satisfactorily in highly controlled operating environments such as border control or immigration check points, under constrained illumination, pose and facial expressions. The performance of these systems in public spaces and civilian operating environments is less than satisfactory, and has several weaknesses and limitations. For example, variations in viewing conditions such as pose and lighting can cause far more changes in facial appearance than variation between different people, resulting in failure to identify a person correctly in real video surveillance footage.

The reason for this could be that the models and algorithms for building these automated surveillance systems were drawn from ill-understood and imperfect models of human visual system functions. Several studies have indicated that the ability of the human vision system is far from perfect, and its ability to recognize unfamiliar faces based on individual facial images in isolation is inherently poor. However, using high level contextual information, humans can identify the person using visual cues much better than other species. This suggests that there is an information fusion from multiple information sources which equips humans with better recognition capabilities as compared to any other species. Likewise for automatic person recognition systems, if the high level contextual information could be inferred by combining multiple heterogeneous sources of weak and non-dominant information sources - for example, the side-views of the face, the facial gestures, limb movements, or the walking gait pattern, which are least discriminative and on their own cannot be used for establishing the identity of the person, but can perhaps provide a unique form of non-intrusive contextual information - it could be possible to enhance the performance of automatic identity recognition systems in video surveillance and civilian
application scenarios. The non-dominant secondary information is normally captured without extra burden on the system, and is available as rich multimedia synchronized data from the same CCTV images as the primary facial data.

Researchers in visual image based person identity recognition have so far focused on the dominant information based on salient and dominant biometric identifiers such as frontal face images, iris, palm prints and fingerprints, and tried to build single-mode and multiple-mode fusion models combining some of these primary identifiers (from face, iris, fingerprint). There was little research effort to include the weak, non-dominant information that gets captured simultaneously in the identification process. For example, it is possible to interpret high level semantic information such as gender, age, ethnicity, emotion and aggression from the same video footage collected for facial video data. If this high level secondary information (though it is a weak biometric) can be used in the decision making process, along with the dominant primary information, it is possible to identify the person and the person’s identity, and the specific actions and intentions with a high level of confidence, reliability and accuracy. Further, using weak secondary biometric identifiers in conjunction with primary biometric identifiers in an appropriate authentication protocol, it is possible to address different levels of user and security requirements based on the application scenario and the operating environment. For instance, for the public and civilian customer access application scenarios, the systems can operate with low and medium level security, but can be made more user friendly and less annoying to use with less false reject rates, whereas for high security application scenarios such as access to high security facility or border control and immigration check points, the system can work at high level of security and accuracy with less false acceptance rates.
This thesis endeavors to address some of these shortcomings in current biometric identification technologies, and makes several contributions towards this goal. Each of these contributions have been peer reviewed by scientific/research community and feedback from the research community was considered to improve the approaches and techniques in each Chapter of this thesis.

The key contributions include:

1. Development of a novel face-gait fusion approach and its validation on a benchmark publicly available human action recognition database (Chapter 3).
2. Extension of the face-gait fusion approach to multi view multimodal context, and its validation to another publicly available benchmark CASIA database with more complex data (Chapter 4).
3. Development of in-house University of Canberra Multimodal Gait (UCMG) Database, addressing the current shortage of high quality data in this area (Chapter 5).
4. Development of a novel face-gait-ear fusion approach and its evaluation with the UCMG database (Chapter 6).

These contributions are described in details in the thesis Chapters and are organized as follows. Next Chapter presents a review of literature and sets the background to the proposed research. Chapter 3 presents the novel face-gait fusion approach with benchmark Swedish human action recognition database. The extension of the this approach for multi view multimodal scenarios is described in Chapter 4, and the evaluation of the improvements achieved is validated with a more complex publicly available benchmark database called CASIA Gait database. This extensive experimental work in these two Chapters has revealed the shortcomings of the current, existing databases, in terms of their capability to emulate real world operating scenarios in video surveillance, leading to development of a new UCMG
database, described in Chapter 5. The development of a novel face-gait-ear fusion approach, a more superior and suitable approach for real world surveillance contexts and its experimental evaluation with UCMG database is presented in Chapter 6. Due to ever-growing need for complex surveillance requirements, it is never possible to develop an absolute, perfect approach that can address all requirements. Hence, the conclusions from the contributions made in this thesis, and possible improvements for further research have been identified and presented in Chapter 7, with some key references used for this research detailed in References Section. The peer-reviewed publications coming out of this research are as shown below:

LIST of PUBLICATION MADE:


9. E. Hossain, G. Chetty, Combination of Physiological and Behavioral Biometric for Human Identification, 8th International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM 2012), Berlin, Germany, July 2012. (Peer reviewed)


15. E. Hossain, G. Chetty, Cross-Camera Feature Level Fusion for Person Identification in Surveillance Videos, 20th International Conference on Neural Information Processing (ICONIP-2013), Seoul, South Korea, November 2013. (Peer reviewed)


BOOK CHAPTER

PATENTS

AusPatent no: 2011101355 “Biometric person identity verification base on face and gait fusion” Hossain, S. M. Emdad; Chetty, Girija

    previousQuery=Chetty&queryString=Chetty&includeAbstractText=on&resultsPerPage=10

AUDIO-VISUAL RECORDINGS

Chapter 2: Literature Review

2.1 Introduction

Prior work of face-gait fusion is not remarkable. Individually has shown bit of effort. Like; face biometric has been used for face-based identity verification system, and gait biometric used for gait analysis on regards to disease particularly, for treating gait disabilities in children and old age people and to assist in early detection of Parkinson’s & other diseases. However, number of research on gait-based identity system still going on. That is based in image procession to detect the human silhouette and associated spatiotemporal attributes. Gait can be affected by several factors including choice footwear, the walking surface, clothing etc. gait recognition system however still in developing stage [2]. Likewise facial analysis used for medical treatment for any facial diseases like; aesthetic treatment [3]

2.2 Physiological Approach

Physiological approach basically mean for facial recognition system or facial analysis. Facial biometric is; in fact physiological trait and facial analysis can be done on facial treatment, recognition or identity verification. More frequently it uses for facial beauty analysis [4]. The beauty analysis is a comprehensive knowledge of laser systems, details and treatment parameters, appropriate patient selection, preoperative and postoperative care, and application of new technologies which can produce aesthetic results that are satisfactory to both the patient and the surgeon. Before any intervention, a thorough facial analysis must be undertaken in order to promulgate an appropriate treatment plan. The term resurfacing may eventually be replaced or reserved for very specific types of facial rejuvenation. Ablative, nonablative, and sublative are terms that further stratify modalities of facial "resurfacing [5]. However, face is physiological biometric trait, and I found multiple use of the trait that is treatment, beauty analysis and person identification or verification. Facial analysis for treatment is new phenomenon at all; it has glorious history from early 16th century in the time of medical science invention [6]. And manual analysis of human face started from 1882 with
biometric Bertillon system for human identity verification [7]. It was a blowing invention by Mr. Bertillon to identity a person by using biometric trait. Earlier stage it was manual measurement of physiological (face, iris, fingerprint etc.) and behavioral (gait, voice, signature etc.) trait of a human. Undoubtedly it was the bottom line of biometric identity verification history. Based Bertillon invention, automated biometric identity verification system by using physiological biometric trait came up in around 1960’s and latest in 2001 scientist came up with automated face recognition system [7], and that was the vital contribution of biometric identification history. In the period of technological advancement change is the constant member of change. Means, change always happens as continuous form. This is because of specific and proven reason. This is either limitation in the system or in the applicability. Likewise, researchers absorbed number of weakness of frontal face recognition system. The weakness I highlighted in next section. By considering the weakness I came up with the idea to used side face with clear image of ear instead of frontal face where user cooperation is not mandatory.

2.3 Behavioural Approach

In my project is multimodal project in the sense that I’m face, ear and gait. Face and ear are soft biometric as well as physiological trait. On the other hand, gait (way of walking) is behavioral trait in the sense that is more on human behavior. This behavior (gait) is a matter of analysis especially in the area of medical science as well in the area of identification. However, gait analysis is used in the medical field to recognize walking problems in patients. Those who suffer from paralysis or sports based injuries too are subjected to gait rehabilitation systems. Efficiency of athletes, especially sprinters can be improved by analysing his or her gait characteristics [8]. The recent developments of gait systems are in relation with biometric identification, the use of gait analysis for the implementation of biometric technology has caught attention all over the world. Consider a scenario in walk
from a distance. Another good example is identification of a terrorist on an airport by analysing his body movements and walking style. In general gait analysis showing its potentiality as futuristic solution includes access control, crime investigation etc [8]. Number of consideration has emerged of gait biometric for identity verification systems. Gait is:

- Distance based identification
- Poor illumination
- Difficult to cheat as it is a non-intrusive behavioural trait and
- User involvement is not required

However, it is simple to understand the concept of gait-based biometric systems. Walking patterns of different individuals differ on the basis of repeatable patterns and a number of specific characteristics. In general, most of the individuals have the ability to recognize their friends and relatives in a crowd or from a distance by analysing their walking styles. Eyes and brain of the observer play an important role in this direction. While using gait for biometrics, computers are made the observers to biometrically identify an individual from specific style of walking [8].

2.4 Computational Approach
I figured recognizing faces from surveillance video footage is quite a challenging problem, and it has been the subject of active research for the past several decades because of its many applications in domains such as covert security and context-aware environments. The face’s appeal as a physiological biometric stems from the several advantages that it offers in terms of being nonintrusive, non-invasive, cost-effective, easily accessible (i.e., face data can be conveniently acquired with a few inexpensive cameras) and relatively acceptable to the general public. However, employing the face for recognition also presents some difficulties since the appearance of the face can be altered by intrinsic factors such as age, expression,
facial paraphernalia (facial hair, glasses, cosmetics, etc.), ethnicity, and gender as well as extrinsic ones such as illumination, pose, scale and imaging parameters (e.g., resolution, focus, imaging noise, etc).

Face recognition has been the focus of extensive research for the past three decades [9]. The approaches for this task can be broadly divided into two categories: 1) Feature-based methods [10], [11] which first process the input image to identify and extract distinctive facial features such as the eyes, mouth, nose, etc. as well as other fiducially marks and then compute the geometric relationships among those facial points, thus, reducing the input facial image to a vector of geometric features. Standard statistical pattern recognition techniques are then employed for matching faces using these measurements. 2) Appearance-based (or holistic) methods [12], [13] which attempt to identify faces using global representations, i.e., descriptions based on the entire image rather than on local features of the face. Though face recognition methods traditionally operate on static intensity images, in recent years, much effort has also been directed towards identifying faces from video [14] as well as from other modalities such as 3D [15] and infra-red [16]. But still limitation is there; one simple limitation reported by the author [17] identify a twin pose by using single biometric also a big challenge, and multimodal application has been advised on the occasion.

Recently, few attempts have been expended on combining various biometrics in a bid to improve upon the recognition accuracy of classifiers that are based on a single biometric. Some biometric combinations which have been tested include face, fingerprint and hand geometry [18]; face, fingerprint and speech [19]; face and iris [20]; face and ear [21]; and face and speech [22]. The fusion of face-ear and gait however, did not attract much attention from the research community. This could be due to difficulty in processing and making sense out of them.
Gait is essentially a behavioural biometric, whose utility for human recognition has not been explored much. Similar to the face, gait too, is a visual cue which can be extracted from video and thus, it appears to offer the same advantages and disadvantages that are presented by the face-ear. However, gait offers some additional benefits in that, unlike the face, it can be acquired from a sequence of low resolution images of a person taken from a distance where the subject’s body occupies too few pixels for other biometric traits to be discerned. Moreover, it is more robust to slight variations in viewpoint as compared to other biometrics and cannot be easily disguised without attracting attention. Nevertheless, gait as a biometric has its limitations since, being a behavioural trait, it can be affected and altered by factors such as clothing, footwear, environmental conditions, emotions, fatigue, drunkenness, pregnancy, injury, disease, aging, weight and load. In addition, gait also suffers from the usual problems associated with extracting visual cues from video such as imperfect foreground segmentation of the walking subject from the background scene, and poor imaging conditions.

However, after long review on capabilities and limitations of present or current generation biometric identification technologies [9], [11], [14], [16] and [18], now I can discuss something on next generation biometric technologies that could play a major role in security and authentication applications. According to authors in [19], [7], the expectations of next generation identity verification involve addressing issues related to application requirements, user concern and integration. Some of the suggestions made to address these issues were use of non-intrusive biometric traits, role of soft biometrics or dominant primary and non-dominant secondary identifiers and importance of novel fusion protocols. I can report here some of the work in progress in laboratory in this direction, where I’m investigating face-ear and gait biometrics as potential candidates. Authors [19] also reported on the usability and accuracy of the system to stand for next generation identity verification technology. Since
my proposed methodology is combination non-intrusive physiological, behavioural and soft biometric, also can be taken public surveillance camera, here at this stage I don’t have any doubt on the applicability to the public space. On the other hand, I already proved the accuracy of my proposed methodology in my very first experiment [23], even though, it was only side face and gait. However, I’m expecting that in combination of side face-ear and gait will make the system more reliable and will increase applicability. Moreover, use of visual biometric traits such as face-ear and gait appears promising as they require no user involvement actively in collecting the data, with camera sensors collecting the information automatically.

Conversely, gait being a weak biometric, is more behavioural and on its own cannot be a powerful biometric trait. The fusion of dominant physiological biometric - the facial image patterns with clear image of ear, and a weak behavioral biometric – the gait patterns, can however, be a powerful combination. They are non-intrusive, inexpensive to deploy, require no cooperation from the subjects, and provide abundant data for analysis. However, there is a lack of efficient algorithms and fusion models for processing the combination of near range side face image patterns and medium/long range gait patterns, and make a sensible decision on the identity of the individual - as a civilian or criminal, client or impostor - and detect their actions as benign or harmful. I’m trying to address this problem by proposing several new computational algorithms and fusion models for processing more than two biometric modalities [24]. From the brilliant result of my 1st experiment, I can say; I’m now one step ahead to put my proposed technology as next generation biometric identity verification technology. 8th March 2011, FBI (Federal Bureau of Investigation) of the United States has announced the capability of next generation biometric identification. Incidence or coincidence that, the system also made of fingerprint. The weakness and applicability of the biometric technology based on fingerprint I already discussed [17]. And it’s proved that the
single biometric trait like; fingerprint, iris, ear, face etc. cannot play a vital role on the next generation biometric identity verification technology. Strength and weakness of the mentioned technology already been published.

Furthermore, earliest biometric system that has been invented 1882 “Bertillon System” [17]. It was kind of manual biometric authentication system and I can nominate the system as 1st generation biometric identification verification system. From that till 1960’s all most all the biometric authentication system was based on ‘Bertillon System”. In the early 1960’s automated biometric identification has taken place to world security era. From 1960 til now research, development and argument has been running equally. Number of good system already developed and implemented. This period I can nominate as 2nd generation of biometric technology or current biometric technology. Even though, lot of good project has been done and implemented successfully, still huger going to find out better, robust, and publically applicable biometric technology to identify a person accurately. The need for better biometric identification trait is clarified by the world strongest law enforcement authority FBI (Federal Bureau of Investigation). It says; the FBI Biometric Center of Excellence (BCOE) will be leveraging the potential of newly emerging biometric technology to allow federal government agencies to increase their identity management capabilities. The BCOE will assist in implementing newly-developed biometric modalities such as facial recognition, iris recognition, and palm print matching into large-scale federal government biometric systems. Research will be performed to support the multimodal fusion of numerous biometrics to result in a significantly more accurate and comprehensive identity management system. The BCOE will also work on developing and enhancing other potential new biometric technologies including footprint and hand geometry, gait recognition, etc [25]. Only single biometric trait, single biometric trait adoption with advance technology, and then multimodal biometric trait has been tested. But today, researchers are agree in one point that,
we have to have multimodal biometric trait. And that can be non-intrusive, physiological, behavioural and strong privacy protected trait [26].

In recent year researcher has been tasted single biometric trait in multimodal approach. Basically they used face to identify a people. Within the face they applied two modal that are facial appearance and facial expression features [27]. By applying principle component analysis (PCA) and linier discriminant analysis (LDA) to the mentioned modal the best result was 99.8% (detection rate). Limitations of face recognition have been deployed earlier in this proposal. And this cannot be a multimodal biometric approach; I may call it as multiple approach of single biometric trait. Another visionary approach on multimodal biometric trait is; fusion on face, ear and gait [28]. Researchers used Gabor and PCA on their match score level and decision level fusion. They used 120 set of data for three (3) different traits. And the best recognition performance that their proposed method achieved is 97.5%. Once again, they applied frontal face along with soft biometric ear that has number weakness on application and implementation. Only side face-ear and gait could have given better result. Still limitations are on fusion level. Researchers could have used PCA and LDA or LLDA on regards to feature fusion level instead of Gabor and PCA.

Finally, my proposed biometric is combination of physiological biometric trait (side face-ear) and behavioural biometric trait (gait). That is absolutely non-intrusive, and strong privacy protected trait [26]. Based on my evolution of number of excellent papers and journals and based on my 1st experiment on side face and gait feature level fusion, I can recap that my proposed multimodal biometric trait going to hold the place of next generation biometric identity verification technology and I’m quite confident here at this stage of my project.
2.5 Combined Approach
As we have seen the ability and potentiality of the tree biometric trait face, ear and gait. I discussed the physiological soft biometric (face and ear) from health prospect as well as from identity prospect. That shows computationally they are strong enough to identity a person. Likewise, gait biometric itself proved the potentially to be a strong biometric trait for identification system. However, researchers already evaluated and implemented number of solution by using single and soft biometric like; face, ear, iris, fingerprint etc. some of them applied few modal together like; ear and face, fingerprint and face, iris and face etc. combination of ear, side face and gait however does not attract much attention from research community in this area. My intention was to develop such system by suing multiple biometric traits which will be robust, reliable, unique and applicable to public surveillance. After my long literature review I have some consideration on trait selection that are;

- The face is a short-range biometric, which can be used effectively for identification only when the subject is close enough to the camera for sufficient details of subject’s facial features to be captured.
- Ear is also short range biometric. But, for facial recognition; user cooperation is needed. Whereas for ear object; user cooperation is not required especially in public surveillance, if it is visible.
- Gait, on the other hand, is a medium to long-range biometric, which can be extracted reliably even from low-resolution imagery and is more invariant to slight changes in viewpoint. Researchers in [2] suggested finding invariant representation from inherently varying biometric signal (profile/side face and gait for example), by using an appropriate digital representation, such that the trait can be recognized despite changes in pore, illumination expression, aging and so on [2].
- Using these three biometric traits together I thought; it would arguably make the
system more robust to variations in subject to camera distance. Also, ear, face and gait are visual cues; both can be extracted from the same modality, (i.e., image sequences of people) precluding the need for separate or specialized equipment.

- Exploiting gait based identification technologies can also benefit health care sector, particularly, for treating gait disabilities in children and old age people those who suffer from paralysis or sports based injuries to be subjected to gait rehabilitation system. Efficiency of athletes, especially sprinters can be improved by analysing his/her gait characteristics. However, gait technology still in developing stage, no model has, as of yet, been developed that is sufficiently accurate and marketable [8].

2.6 Summary
The face, ear and gait biometrics make use of apparently independent personal characteristics: face recognition systems exploit the relatively detailed appearance of the facial surface, while gait recognition methods capture data from the coarse body shape as it changes over time. Consequently, some conditions that sharply degrade the performance of face recognition systems, such as large variations in illumination and facial expressions, affect gait to a much lesser extent or not at all. Similarly, some conditions that adversely affect the accuracy of gait recognition, such as clothing, footwear, and load, do not influence the performance of face recognition systems. Therefore, it was reasonable to believe that combining these complementary cues would improve the recognition accuracy. And that’s what proved by developing my project up to this stage.
Chapter 3: Face-Gait Fusion with Human Action Recognition Dataset

3.1 Introduction

Biometric person identification is a common technological tool for identity verification. It carries significant importance for national or international security. All most each and every part of human body is unique; some of the significant ones have been used for developing automate identity verification systems. Fingerprint, palm print, face, iris, ear [7], [29] etc. have been used immensely for current generation of person authentication technologies. There are still challenges in this area, and need for better biometric modalities, development of novel approaches and techniques are being an ongoing process. Video surveillance in public places and facilities has become omnipresent, and has become the first line of defence for protecting assets and people for different types of operating scenarios and applications – be it a civilian public space for access control to a facility, or financial and transaction oriented applications, or the high security immigration and border control check points. It has become an enabler of trust, integrity and security in the new Digital Economy. The need for non-intrusive biometric modalities enjoys significant user acceptability. Though any one biometric modality on its own cannot address all the challenges, and importance of combining the information from multiple biometric modalities holds significant promise. Having reviewed the capabilities and limitations of present current generation biometric identification technologies [7], [30], [31], [32], [33] and [34], we now discuss some of the next generation biometric technologies that could play a major role in security and authentication applications. According to authors in [7], the expectations of next generation identity verification involve addressing issues related to application requirements, user concern and integration. Some of the suggestions made to address these issues were use of non-intrusive biometric traits, role of soft biometrics or dominant primary and non-dominant secondary identifiers and importance of novel fusion protocols. I reported here some of the work in progress in our laboratory in this direction, where I have been investigating face and
gait biometrics as potential candidates. Use of visual biometric traits such as face and gait appears promising as they require no user involvement actively in collecting the data, with camera sensors collecting the information automatically. However, gait being a weak biometric, is more behavioural and on its own cannot be a powerful biometric trait. The fusion of dominant physiological biometric - the facial image patterns, and a weak behavioral biometric – the gait patterns, can however, be a powerful combination. Both are non-intrusive, inexpensive to deploy, require no cooperation from the subjects, and provide abundant data for analysis. However, there is a lack of efficient algorithms and fusion models for processing the combination of near range face image patterns and medium/long range gait patterns, and make a sensible decision on the identity of the individual - as a civilian or criminal, client or impostor - and detect their actions as benign or harmful.

3.2 Background
Person identity verification from arbitrary views in low resolution surveillance video is a very challenging problem, especially when one is walking at a distance. Over the last few years, recognizing identity from gait patterns has become a popular area of research in biometrics and computer vision, and one of the most successful applications of image analysis and understanding. Gait recognition is one of the new and important biometric technologies based on behavioural characteristics, and it involves identifying individuals by their walking patterns. Gait can be captured at a distance by using low resolution devices, while other biometrics needs higher resolution. Also, gait is difficult to disguise, and can be performed at a distance or at low resolution, and requires no body-invading equipment to capture gait information. Gait recognition can hence be considered as a next-generation identity verification technology, with applicability to many civilian and high security environments such as airports, banks, military bases, car parks, railway stations etc. Further, gait is an inherently multimodal biometric as proposed by Murray et al in [35], suggesting that there
are 24 different components to human gait, and involves not just the lower body or legs, but also the upper body in terms of motion associated with the torso, the head and the hands. If all gait movements from full body images can be captured, it can be a truly a unique biometric.

In this research I proposed a multimodal fusion technique by combining face and gait features in learning subspaces based on principal component analysis (PCA) and linear discriminant analysis (LDA). Further, by processing the fusion features with multivariate Gaussian classifiers, it is possible to capture several inherent multimodal components present in human Gait. Extensive experiments conducted on a publicly available gait database [36] suggest that to obtain optimal performance, an integrated face, body and gait cues obtained from video sequences and processed appropriately with learning approaches mentioned above, can result in a simple, practical and robust identity verification technique in spite of poor quality data from surveillance video with significant degradations in operating environments. The recognition of people is of great importance, since it allows us to have a greater control about when a person has access to certain information, area or simply to identify if the person is the one who claims to be [2] [37]. And one natural tool to identify a person is the biometric trait. Automated face recognition technology [2], [37], [38] first captured the public attention from the media reaction to a trial implementation at the January 2001 super bowl, which captured surveillance image and compared them to a database mug shots [38]. From 1960s till now vast number of research works have been conducted on biometric person authentication. Several research articles have been reported involving the use of signatures, fingerprints, face and voice information [39]. For face recognition systems, the performance of 2D face matching systems depends on capability of being insensitive of critical factors such as facial expression, makeup and aging, but also relies upon extrinsic factors such as illumination difference, camera viewpoint, and scene geometry [40]. Further, the 2D face recognition
systems are vulnerable to pose, and illumination variations. Use of 3D face can make systems robust to pose and illumination variations. The state of the art 3D face recognition technique using isogeodesic stripes was proposed in [40], However, progress in 3D identification approaches has been slow as it suffers from higher infrastructure costs and limited availability of databases [41]. Hence, alternate biometric traits and combination of different types of biometric traits was explored by the biometric security community. Subsequently, due to increase in global demand for automated security and surveillance products, there was a proliferation of research works on identity verification based on different biometric modalities [7], [30], and [31]. Several research works have also reported importance of using multiple modalities instead of single biometric trait in order to enhance the accuracy and robustness [30], [31] and [32]. However, most of the systems have been tested in controlled laboratory environments, and it is a huge challenge to achieve similar accuracies and robustness in real world public surveillance applications. Further, the current generation of identity authentication systems are based on modalities based on fingerprint, palm print, face, iris, ear biometric traits [33], [34] and [29]. These modalities are of limited use for deployment in public surveillance scenarios or performing authentication at a distance. Lately, an increasing need for surveillance in public places and facilities from a distance has been felt due to terrorist attacks or attack on public assets. And the automated video surveillance systems serve as the first line of defence for protecting assets and people for different types of operating scenarios and applications – be it a civilian public space for access control to a facility, or financial and transaction oriented applications, or the high security immigration and border control check points. It has become an enabler of trust, integrity and security in the new Digital Economy [42], [43] and [24]. However, the surveillance systems designed to work in high security environments fail miserably when deployed for day-to-day civilian environments, due to unconstrained noisy and non-ideal
operating conditions of civilian environments. Integrating multiple sources of information can solve some of the problems with these systems; and this integration could be at a lower level involving sensors, data and feature extractors, or at a higher level, at decision or at a score level.; or it could also be at an ancillary or side level, consisting of higher level context information. Humans have long been using such multiple levels of information sources as cues to perform any identification tasks, particularly in difficult scenarios. This then suggests that there is an information fusion from multiple information sources which equips humans with better recognition capabilities as compared to any other species. Likewise for automatic person recognition systems, if multiple heterogeneous sources of information could be combined and used – for example, the side-views of the face, the facial gestures, limb movements, or the walking gait pattern, which may be least discriminative and on their own cannot be used for establishing the identity of the person, but can perhaps provide a unique form of contextual information - it could be possible to enhance the performance of automatic identity recognition task in video surveillance systems operating in unconstrained civilian operating environments. The non-dominant secondary information is normally captured without extra burden on the system, and is available as rich multimedia synchronized data from the same CCTV images as the primary facial.

The proposed multimodal approach in this chapter is based on side face and gait, and both can be extracted from low resolution imagery. And it is not necessary to have clear face or gait, making it suitable to collect data irrespective of user’s disable, aged or any gender. The potential of gait as a powerful biometric has been explored in some of the recent works [24], [17], though inherent multimodal components present in the whole body during walking has not been much exploited by the research community. In this experiment I investigated the potential of combining rich multimodal information available form face body (torso) and gait biometric cues, to ascertain the human identity in low resolution surveillance videos with
unconstrained operating environments. Face and gait based identity verification serves another important purpose – and that is of addressing sensitive privacy issues associated with capture and storage of biometric data. Some of the most important challenges for diffusion of biometrics in day-to-day civilian applications are issues related to invasion of privacy. In [26], an extensive study has shown that physiological biometrics as having no negative impact whatsoever on the privacy. That is an excellent motivation for us to investigate gender-specific face, body(torso) and gait cues during walking as a powerful biometric with inherent multimodality for ascertaining the identity of a person. Further, these video based cues can be captured remotely from a distance, and by using an appropriate biometric identification protocol as suggested by authors in [44], it can be ensured that sensitive privacy concerns are addressed as well. An appropriate protocol as in [44] can ensure that the identification system is not misused and that function creep (i.e. use for another purpose is prevented). This means in particular that a component should not be able to learn more information than what is really needed for a correct result. In fact our proposed fusion of face and gait cues captured from low resolution surveillance videos (“security check: pass”) needs strong algorithms and processing techniques to be of any use for establishing identity, and of no use without them, and hence can safe-guard the privacy to a large extent automatically.

3.3 Methodology
For experimental evaluation of our proposed face and gait fusion scheme, I used a publicly available video database of human actions [36]. This video database contains six types of human actions (walking, jogging, running, boxing, hand waving and hand clapping) performed several times by 25 subjects in four different scenarios: outdoors s1, outdoors with scale variation s2, outdoors with different clothes s3 and indoors s4. Currently the database contains 2391 sequences.
All sequences were taken over homogeneous backgrounds with a static camera with 25fps frame rate. The sequences were down-sampled to the spatial resolution of 160 × 120 pixels and have a length of four seconds in average. I used only the walking sequences for our experiments. Figure 1 shows some of the sample images from the walking video sequences, and Figure 2 shows multiple frames of the sequences for a person walking in the video clip.

For all our experiments I used 100 video sequences for 25 people. There were 19 males and 6 females in the entire walking dataset. I performed some image pre-processing steps corresponding to cropping, filtering and histogram equalization and then extracted features based on PCA (principal component analysis) and LDA (linear discriminant analysis). I used separate set for performing training and testing. The low dimensional PCA and LDA features
were then classified by a Bayesian classifier. I examined four different classifiers, the nearest neighbour \((k\text{-}NN)\), the Bayesian linear and the Bayesian quadratic classifiers, and the Mahalanobis classifier. The combination of the low dimensional, discriminative PCA and LDA features along with powerful multivariate Bayesian linear/quadratic classifiers allow us to achieve significant improvement in recognition accuracy as compared to conventional Euclidean distance based methods reported predominantly in previous works. This is because Bayesian classifiers have the flexibility to incorporate prior information, and can predict how a system’s performance will change when there is a mismatch in train and test conditions. And \(k\text{-}NN\) is very effective simple classifier with noise reduction capabilities [45], [46]. The schematic for the proposed multimodal identification scheme is shown in Figure 3. A brief description of PCA and LDA feature processing technique is given next.

Figure 3: Schematic for proposed multimodal identification scheme based on fusion of side face and gait cues extracted from low-resolution video
3.4 Motivation for Face-Gait Fusion
Since the performance of any classifier is more sensitive to some factors and relatively
invariant to others, a recent trend has been to combine individual classifiers in order to
integrate their complementary information and, therefore, create a system that is more robust
than any individual classifier to variables that complicate the recognition task. In [44],
researchers have shown that integrating multiple biometrics does indeed result in consistent
performance improvement while in [47], authors have empirically demonstrated that as the
number of classifiers combined increases, so does the recognition accuracy. The encouraging
results reported in the literature for such systems in conjunction with the conclusions reached
by the previous studies mentioned in the background section of this chapter, provide a strong
basis for believing that a system constructed by combining various biometric characteristics
is going to yield better recognition rates than the individual classifiers for those traits.
Further, using multiple biometrics is a viable solution to real-world problems, such as non-
universality of some biometric traits (e.g., some people’s fingerprints cannot be reliably
extracted because of the poor quality of the ridges also possible to do fingerprint alteration
[42], unavailability of data for a certain biometric (e.g., visual cues such as face, ear, etc.
might be occluded in surveillance videos) and criminal activity (i.e., attempts to fool the
single-biometric based system by duplicating the biometric trait or breaching the system). In
light of the above, some specific reasons for considering investigation of face and gait
biometric fusion are as follows:
➢ The face is a short-range biometric, which can be used effectively for identification
only when the subject is close enough to the camera for sufficient details of subject’s
facial features to be captured.
➢ Gait, on the other hand, is a medium to long-range biometric, which can be extracted
reliably even from low-resolution imagery and is more invariant to slight changes in
viewpoint. Researchers in [2] suggested finding invariant representation from
inherently varying biometric signal (profile/side face and gait for example), by using
an appropriate digital representation, such that the trait can be recognized despite
changes in pore, illumination expression and aging [2].

- Using these two biometric traits together would arguably make the system more
  robust to variations in subject to camera distance. Also, both face and gait are visual
cues; both can be extracted from the same modality, (i.e., image sequences of people)
precluding the need for separate or specialized equipment.

- Further, the face and gait biometrics make use of apparently independent personal
  characteristics: face recognition systems exploit the relatively detailed appearance of
  the facial surface, while gait recognition methods capture data from the coarse body
  shape as it changes over time. Consequently, some conditions that sharply degrade the
  performance of face recognition systems, such as large variations in illumination and
  facial expressions, affect gait to a much lesser extent or not at all. Similarly, some
  conditions that adversely affect the accuracy of gait recognition, such as clothing,
  footwear, and load, do not influence the performance of face recognition systems.
  Therefore, it is reasonable to believe that combining these complementary cues would
  improve the recognition accuracy.

### 3.5 Principle Component Analysis (PCA)

PCA is a useful statistical technique that has found application in fields such as face
recognition and image compression, and is a common technique for finding patterns in data
of high dimension [45], [46]. Let \{x_1, x_2, \ldots, x_n\}, x_k \in \mathbb{R}^N, be \(n\) random vector, where \(N\) is
the dimensionality of the vector obtained by concatenation of an image row-by-row. The
covariance matrix is defined as \(\Sigma_x = E[(x - E(x))(x - E(x))^T]\), where \(E(\cdot)\) is the expectation
operator and \(T\) denotes the transpose operation. The covariance matrix \(\Sigma_x\) can be factorized
into the following form:
\[ \Sigma x = \Phi \Lambda \Phi \] 

where \( \Phi = [\Phi_1, \Phi_2 \ldots \Phi_N] \in \mathbb{R}^{N \times N} \) is the orthogonal eigenvector matrix of \( \Sigma x \); \( \Lambda = [\Lambda_1 \Lambda_2 \ldots \Lambda_N] \in \mathbb{R}^{N \times N} \) is the diagonal Eigen value matrix of \( \Sigma x \) with diagonal elements in descending order [45], [46]. One important property of PCA is its optimal signal reconstruction in the sense of minimum mean square error (MSE) when only a subset of principal components are used to represent the original signal. An immediate application of this property is the dimensionality reduction [45], [46].

\[ y_k = \mathbf{P}^\text{PC1}[x_k - E(x)], \quad k = 1, 2, \ldots, n, \quad \ldots \ldots \quad (2) \]

where \( \mathbf{P}_{\text{PC1}} = [\Phi_1, \Phi_2 \ldots \Phi_m], m \leq N \). The lower dimensional vector \( y_k \in \mathbb{R}^m \) captures the most expressive features of the original data \( x_k \) [26], [48], [46].

If \( f \in \mathbb{R}^{N1} \) and \( g \in \mathbb{R}^{N2} \) represent the PCA vectors corresponding to a person video data, where \( N1 \) and \( N2 \) are the dimensionality of the face and the gait feature spaces, respectively. I obtain low dimensional feature vectors, \( f' = \mathbf{M}_f f \) and \( g' = \mathbf{M}_g g \), by using the PCA method as in Eq. (2). \( \mathbf{M}_f \) and \( \mathbf{M}_g \) are the PCA transformation matrices for face and gait, respectively. I choose a subset of principal components to derive the lower dimensional face and gait features, \( f' \in \mathbb{R}^{m1} \) and \( g' \in \mathbb{R}^{m2} \), where \( m1 \) and \( m2 \) are the dimensionality of the reduced face feature space and gait feature space, respectively. On one hand, I hope to lose as little representative information of the original data as possible in the transformation from the high dimensional space to the low dimensional one. On the other hand, the eigenvectors corresponding to the small Eigen values are excluded from the reduced space so that I can obtain more robust PCA & LDA projection as well as reduce the problem of curse of dimensionality [45], [46]. The Eigen value spectrum of the covariance matrix of the training data supplies useful information regarding the choice for the dimensionality of the feature space. Before face features and gait features are normalized to have their values lie within similar ranges. I used a linear method [45], [46] which provides normalization via the
respective estimates of the mean and variance. For the \( j \)th feature value in the \( i \)th feature vector \( w_{ij} \), we have:

\[
\hat{w}_{ij} = \frac{(w_{ij} - w_j')}{\delta_j},
\]

\( i = 1, 2, \ldots, I \),

\( j = 1, 2, \ldots, L \), ………………………………(3)

where \( j w = (1/I) \sum_{i=1}^{I} w_{ij} \) and \( S_B \). \( I \) is the number of available feature vectors and \( L \) is the number of features for each feature vector. The resulting normalized features have zero mean and unit variance. To take advantage of information for a walking person in video, I use all possible combinations of complete images, side face features and gait features to generate the maximum number of vectors \( h \). specifically; I have two feature vectors of side face and two feature vectors of gait for one person from one video. Therefore, we have four concatenated features \( h \) for one person from one video. Generation of all possible low dimension vectors \( h \) from PCA analysis for side face and gait data helps to reduce the problem of curse of dimensionality for the subsequent LDA transformation [45], [46].

3.6 Linear Discriminant Analysis (LDA)

Suppose that \( w_1, w_2, \ldots, w_c \) and \( n_1, n_2, \ldots, n_c \) denote the classes and the number of concatenated feature vectors \( h \) within each class, respectively, with \( w = w_1 \cup w_2 \cup \cdots \cup w_c \) and \( n^\wedge = n_1 + n_2 + \cdots + n_c \). Note that the value of \( n^\wedge \) is two times of \( n_c \) is the number of classes. LDA seeks a transformation matrix \( W \) that maximizes the ratio of the between-class scatter matrix \( S_B \) to the within-class scatter matrix \( S_w \)

\[
\sum_{i=1}^{c} n_i(\bar{M}_i - \bar{M})(\bar{M}_i - \bar{M})^T \……………………………………………….(4)
\]

\[
S_w = \frac{J(w)|WTS_B W'|}{|WTS_w W|}
\]

The within-class scatter matrix is

\[
S_w = \sum_{i=1}^{c} \sum_{h_{i,c}} (h - \bar{M}_i)(h - \bar{M}_i)^T
\]
and the between-class scatter matrix is

\[ S_B = \sum_{i=1}^{c} n_i (M_i - M)(M_i - M)^T \]  

(5)

where \( M_i = M_s = (1/n_i) \sum h \varepsilon_{hi} \)

and \( M = (1/n^*) \sum h \varepsilon \)

are the means of the class \( i \) and the grand mean, respectively.

I use all possible combinations of side face features and gait features to generate the maximum number of concatenated feature vectors based on the characteristics of face and gait. Specifically, four concatenated features are constructed based on two face features and two gait features for one person from each video. Let \( V_i, \ i = 1, 2, \ldots, c \), the mean of the training synthetic features of class \( i \), be the prototype of class \( i \). The unknown person is classier to class \( K \) to whom the synthetic feature \( p \) is the nearest neighbour[26]:

\[ ||p - V_K|| = \min ||p - V_i|| \]  

(6)

When multiple synthetic features are obtained for one person, Eq. (6) means that the unknown person is classified to the class which has the minimum distance out of all the distances corresponding to all the classes [26]. Instead of using traditional Euclidean distance based classifiers I applied learning classifiers (Bayesian linear and quadratic classifiers). The Bayesian linear and quadratic discriminant classifier uses Bayesian decision rule for classifying a set of learned feature vectors to a class [46]. While the linear classifier fits a multivariate normal density to each group, with a pooled estimate of covariance, the quadratic discriminant classifier fits MVN (multivariate normal) densities with covariance estimates stratified by group. Both methods use likelihood ratios to assign observations to groups. Given a set of classes \( M \) characterized by a set of known parameters in model \( \Omega \) a set of extracted feature vector \( X \) belongs to the class which has the highest probability. This is shown in Eq.(7)) and is known as Bayes decision rule.
To calculate the \( a \)-posteriori probability shown, I used Bayes law of statistics which finally by assuming that features are distributed normally, leads to a quadratic classifier format known as Bayes Quadratic classifier [46]. The model \( \Omega \) consists of the mean and the covariance of our training vectors, and likelihoods are calculated as stated above. The details of the experiments carried out in the next Section.

### 3.7 Experiments

I performed different sets of experiments for examining the discriminating ability of proposed feature extraction transformation and classifier techniques. Initial experiments involved creating separate gender-specific datasets and examining the performance of each experiment. However, the performance was relatively poor due to imbalance in data available in number of females in the data set. Hence they are reported in this chapter. For each experiment I used datasets corresponding to face-only partial gait (lower body) and full gait (full images) for all 25 people for examining the performance of single mode and fusion of PCA and LDA features for different types of classifiers.

#### 3.7.1 Feature Level Fusion

In feature-level fusion, the feature sets originating from multiple biometric sources are consolidated into a single feature set by the application of appropriate feature normalization, transformation, and reduction schemes. The primary benefit of feature-level fusion is the detection of correlated feature values generated by different biometric algorithms thereby identifying a compact set of salient features that can improve recognition accuracy. Eliciting this feature set typically requires the use of dimensionality reduction methods and, therefore, feature-level fusion assumes the availability of a large number of training data [49].
3.7.1.1 Recognition Performance with PCA-Features

For the first set of experiments I applied PCA transformation and performed classification with Bayesian (linear/quadratic) and k-nearest neighbour classifiers. Table 1 shows the recognition accuracies achieved for PCA only features. For this experimental scenario, I received 80% recognition accuracy for Bayesian-linear classifier, 90% accuracy for Bayesian quadratic and 1-NN classifier. Though I expected a 100% accuracy for face-only mode, what I found was that quality of side face images was very poor, resulting in failure to recognize some poor quality faces. However, PCA was still able to model the low resolution side faces pretty good. This reduction of recognition accuracy is expected as PCA cannot capture the dynamic gait variations accurately no matter how efficient classifier is, The recognition performance (shows in Figure 4 and 5) gets worst for partial gait images (40 - 60%) and by including full gait images, it is slightly better (60-70%), perhaps because of inclusion of torso in full images. (The persons wore same clothes in train and test sessions.)

Figure 4: Eigen face of side-face only
Next, I performed experiments with fusion of face and gait (both partial and full gait images), and recognition accuracies achieved is shown in Table 2. An improvement in recognition performance was achieved with face-partial gait fusion resulting an accuracy of 70% for all three classifiers, and face-full gait fusion resulting in further improvement of accuracies (60 – 90%), with multivariate Bayesian Classifier performing the best with an accuracy of 90%.

<table>
<thead>
<tr>
<th>Name</th>
<th>Face only</th>
<th>Partial Gait</th>
<th>Full Gait</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-NN Classifier</td>
<td>90%</td>
<td>55%</td>
<td>70%</td>
</tr>
<tr>
<td>Bayesian-Linear</td>
<td>80%</td>
<td>40%</td>
<td>60%</td>
</tr>
<tr>
<td>Bayesian-Quadratic</td>
<td>90%</td>
<td>65%</td>
<td>65%</td>
</tr>
</tbody>
</table>
Table 2: PCA with face-gait fusion with Bayesian Classifiers and 1-Nearest Neighbour Classifier

<table>
<thead>
<tr>
<th>Name</th>
<th>Face-Partial Gait</th>
<th>Face-Full Gait</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-NN Classifier</td>
<td>70%</td>
<td>85%</td>
</tr>
<tr>
<td>Bayesian-Linear</td>
<td>70%</td>
<td>60%</td>
</tr>
<tr>
<td>Bayesian-Quadratic</td>
<td>70%</td>
<td>90%</td>
</tr>
</tbody>
</table>

3.7.1.2 Recognition Accuracies with PCA-LDA Features

For this set of experiments, I transformed the PCA vectors in LDA space, and there was a significant improvement in recognition without fusion and with fusion. The results without fusion is shown in Table 3, and as can be seen from this table, even the gait only modes (both partial and full gaits) resulted in good accuracies (80% – 95%), with Bayesian quadratic classifier performing the best, perhaps due to its ability to model the nonlinear gait dynamics accurately. When I performed the face and gait fusion of LDA transformed features, I got a remarkable improvement in accuracies with all three types of classifiers resulting in 100% accuracy. Thus a combination of PCA-LDA processing along with efficient classifiers, it was possible to identify a walking human from a distance even in low resolution video with poor backgrounds. Further, for all modes the multivariate classifier, particularly the quadratic one performs better as compared to 1-NN classifier used by several earlier reported studies. Also, I found the LDA has a remarkable capability to model the gait variations in the person and retain the identity specific information. Figure 4 shows the first 8 most significant Eigen Images of faces, partial gaits and full gaits and Figure 5 shows that most significant Fisher Images of faces, partial gaits and full gaits.

Table 3: LDA with Bayesian Classifiers and 1-Nearest Neighbour Classifier

<table>
<thead>
<tr>
<th>Name</th>
<th>Face-Only</th>
<th>Partial Gait</th>
<th>Full-Gait</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-NN Classifier</td>
<td>95%</td>
<td>80%</td>
<td>85%</td>
</tr>
<tr>
<td>Bayesian-Linear</td>
<td>95%</td>
<td>90%</td>
<td>90%</td>
</tr>
<tr>
<td>Bayesian-Quadratic</td>
<td>95%</td>
<td>95%</td>
<td>90%</td>
</tr>
</tbody>
</table>
3.7.2 Score Level Fusion
I performed different sets of experiments for examining the discriminating ability of proposed feature extraction transformation and classifier techniques. Further I also compared the performance of score and feature-level fusion (schematic shown in Figure 5 and 6). The recognition performance of single mode face and gait features, and with fusion of face and gait features at score-level and at feature-level, are discussed in next few sub-sections.

3.7.2.1 Recognition Performance with PCA-Features
For the first set of experiments I used PCA features for training and testing, with a Bayesian (linear/quadratic) and k-nearest neighbour classifiers for classification. Table 5 shows the recognition accuracies achieved for PCA only features. For this experimental scenario, I received 85% recognition accuracy for Bayesian-linear classifier, 90% accuracy for Bayesian quadratic, and 95% for 1-NN classifier. Though I expected a 100% accuracy for face-only mode, what I found was that quality of side face images was very poor, resulting in failure to recognize some poor quality faces.

Though I expect a 100% accuracy for face-only mode, what I found was that quality of side face images was very poor, resulting in failure to recognize some poor quality faces.

Table 4: LDA face - gait fusion with Bayesian Classifiers and 1-Nearest Neighbour Classifier

<table>
<thead>
<tr>
<th>Name</th>
<th>Face-Partial Gait</th>
<th>Face-Full Gait</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-NN Classifier</td>
<td>95%</td>
<td>100%</td>
</tr>
<tr>
<td>Bayesian-Linear</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Bayesian-Quadratic</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 5: PCA with Bayesian Classifiers and 1-Nearest Neighbour Classifier

<table>
<thead>
<tr>
<th>Name</th>
<th>Face-Only (PCA)</th>
<th>Gait-Only (PCA)</th>
<th>Face-Gait (Feature Fusion)</th>
<th>Face-Gait Score Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-NN Classifier</td>
<td>85%</td>
<td>45%</td>
<td>75%</td>
<td>65%</td>
</tr>
<tr>
<td>Bayesian-Linear</td>
<td>90%</td>
<td>50%</td>
<td>65%</td>
<td>60%</td>
</tr>
<tr>
<td>Bayesian-Quadratic</td>
<td>95%</td>
<td>50%</td>
<td>70%</td>
<td>55%</td>
</tr>
</tbody>
</table>
Next, I performed experiments for gait only mode, and I achieved a poor recognition accuracy of 45% recognition for Bayesian linear classifier, 50% for Bayesian-quadratic classifier and 50% of 1-NN classifier. Once again, PCA features for gait only mode failed badly because of the inability of PCA technique to capture the gait dynamics of each person. However, when I integrated the face-only information with gait information, the performance improved significantly, resulting in an accuracy of 75%, 65% and 70% for Bayesian-linear, Bayesian-quadratic and 1-NN classifiers respectively.

Further, as can be seen in Table 5, feature level fusion performs better than score level fusion for all three classifiers ascertaining the inherent multimodality in face and gait, which is modeled better with feature-level fusion mode as compared to score-level fusion. For all the experiments in this set I used 40 PCA feature dimensions.

3.7.2.2 Recognition Accuracies with PCA-LDA Features
For this set of experiments, I obtained the PCA transformation first and then PCA features were transformed in the LDA space again, training and testing was performed on PCA-LDA vectors, with this, I achieved 100% accuracy for face-only data set. For gait only data set, I achieved a recognition accuracy of 90% for Bayesian-linear, 90% for Bayesian-quadratic, and 80% for 1-NN classifier. Combining the face-gait features in PCA+LDA subspace it was possible to achieve a recognition accuracy of 100% for all three types of classifiers. Figure 7 and 8 shows the Eigen face for both experiments.

Since the face only classifier in PCA-LDA subspace results in 100% accuracy, it would appear that there is no need for fusion with gait features. However, the dimensionality of face only PCA-LDA features was 40 for achieving 100% accuracy, whereas, the dimensionality of features needed to achieve 100% accuracy was much lesser when face and gait features were fused. I needed 20 features for feature-level fusion and 30 features with score-level fusion to achieve 100% accuracy.
As can be seen in Table 6, PCA features in LDA subspace were capable in capturing the person-specific gait variations accurately for all three classifiers. So it was a synergistic fusion, with PCA helpful in reducing the dimensionality and LDA capturing inter-person and intra-person gait associated variations accurately. Another interesting observation was though it is well known in literature, that the score-level fusion results in better performance than feature level fusion, I found that the number of features needed for score fusion is higher (30 as compared to 20 features for feature-level fusion before concatenation). Thus could be
because score level fusion does not preserve the inherent multimodality present in face and gait as well as feature-level fusion can do.

### Table 6: PCA-LDA with Bayesian Classifiers and 1-Nearest Neighbour Classifier

<table>
<thead>
<tr>
<th>Name</th>
<th>Face-Only (PCA-LDA)</th>
<th>Gait-Only (PCA-LDA)</th>
<th>Face-Gait (Feature Fusion)</th>
<th>Face-Gait Score Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-NN Classifier</td>
<td>100%</td>
<td>90%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Bayesian-Linear</td>
<td>100%</td>
<td>90%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Bayesian-Quadratic</td>
<td>100%</td>
<td>80%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

#### 3.7.3 Holistic vs. Hierarchical Fusion

For this fusion approach, I investigated the importance of a new fusion protocol, by integrating face and gait cues for the single camera case. I employed a view invariant gait recognition algorithm for gait recognition. A sequential importance sampling based algorithm used for probabilistic face recognition from video. Then I applied decision fusion to combine the results of our gait recognition algorithm and the face recognition algorithm. Finally I consider two new fusion protocols: hierarchical and holistic. The first protocol involved using the gait recognition algorithm as a filter to pass on a smaller set of candidates to the face recognition algorithm. The second protocol involved combining the similarity scores obtained individually from the face and gait recognition algorithm’s simple rules like; the SUM, MIN and PRODUCT has been used for combining the scores.

For all our experiments I used 100 video sequences for 25 people. There were 19 males and 6 females in the entire walking dataset. I performed some image pre-processing steps corresponding to background segmentation, cropping, filtering and histogram equalization of images of the walking human and then extracted features based on PCA (principal component analysis) and LDA (linear discriminant analysis). I used separate set for performing training and testing. The low dimensional PCA and LDA features were then classified by a Bayesian classifier as described in figure 9.
I examined three different classifiers, the nearest neighbour ($k$-NN), the Bayesian linear and the Bayesian quadratic classifiers. The combination of the low dimensional, discriminative PCA and LDA features along with powerful Bayesian classifiers allow us to achieve significant improvement in recognition accuracy as compared to conventional Euclidean distance based methods reported predominantly in previous works. This is because Bayesian classifiers have the flexibility to incorporate prior information, and can predict how a system’s performance will change when going from one environment to another or when going from one type of testing to another [48]. And $k$-NN is very effective simple classifier with noise reduction capabilities [46]. Further, I examined holistic and hierarchical fusion approach for combining face and gait features. The schematic for the proposed multimodal identification scheme is shown in Figure 5 and 6. The schematic for the proposed multimodal identification scheme is shown in Figure 9. While the fusion strategy/protocol shown in Figure 9 is an holistic fusion approach, which uses a fusion of scores from face and gait classifiers (with equal weightage to face and gait scores), I also examined the hierarchical fusion of face and gait features (shown in Figure 10). In hierarchical fusion mode, the face and gait biometric classifier work in a tandem or cascade mode, where only one classifier
(face/gait) makes an ID accept/reject decision. However if recognition accuracy is low, additional confidence on decision level is obtained with a 2nd biometric classifier. Figure 10 shows a hierarchical fusion scheme with face biometric for a first stage classification, followed by gait classification; Next Section presents a brief description of PCA and LDA feature processing technique.

Figure 10: Hierarchical Fusion scheme with Gait classifier as 1st stage and Face Classifier as a 2nd stage classifier

In this set of experiments, I examined the hierarchical versus holistic fusion. The holistic fusion is essentially same as score level fusion. While the score level fusion in Figure 6 uses 30 features, for comparing the performance of hierarchical vs. holistic fusion, I used 20 features for each. With 20 features, both the classifiers have less than 100% accuracy, and with gait classifier as the first stage classifier, the hierarchical fusion performance is as shown in Table 7.

Table 7: Holistic vs hierarchical Fusion

<table>
<thead>
<tr>
<th>Name</th>
<th>Gait-Only (PCA-LDA)</th>
<th>Face-Only (PCA-LDA)</th>
<th>Face-Gait Hierarchical Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian-Linear</td>
<td>50%</td>
<td>80%</td>
<td>90%</td>
</tr>
<tr>
<td>Bayesian-Quadratic</td>
<td>60%</td>
<td>85%</td>
<td>95%</td>
</tr>
<tr>
<td>1-NN Classifier</td>
<td>50%</td>
<td>80%</td>
<td>85%</td>
</tr>
</tbody>
</table>
I set the threshold for 2nd stage face classifier to be invoked to 95%, so that when classifier recognition accuracy is greater than 95%, the ID accept/reject decision from face classifier score was considered valid. However, when gait classifier accuracy is less than 95%, the gait classifier is also used for making a decision, by fusing the face and gait scores.

Hence, the confidence level of the ID accept/reject decision is enhanced by 2nd stage classifier score with face PCA-LDA features [48].
3.8 Fusion of Ear, Side-face and Gait
In this experiment, I proposed a novel approach of multimodal biometric identify verification system based on side face, ear and gait biometric. It involves feature level fusion of three biometric traits. I applied two different orthogonal transformation approaches that are PCA and LDA. To classify or identify a person from feature level fusion; For classification I used both Bayesian classifiers (linear and quadratic) and kNN classifier. The experimental evaluation of proposed methods for a freely available low resolution video database showed promising outcomes of the methods And I received around 90% correct recognition rate

3.8.1 Experiments and Results:
For experimental evaluation of our proposed multimodal scheme, I used a publicly available video database of human actions [36]. This video database contains six types of human actions (walking, jogging, running, boxing, hand waving and hand clapping) performed several times by 25 subjects in four different scenarios: outdoors s1, outdoors with scale variation s2, outdoors with different clothes s3 and indoors s4. Currently the database contains 2391 sequences. All sequences were taken over homogeneous backgrounds with a static camera with 25fps frame rate. The sequences were down-sampled to the spatial resolution of 160 × 120 pixels and have a length of four seconds in average. I used only the walking sequences for our experiments and Figure 13 shows some of the sample images from the walking video sequences.
However, for our experiments I used 72 video sequences for 18 people. I performed image pre-processing steps corresponding to background segmentation, cropping, filtering and histogram equalization of images of the walking human and then extracted features based on PCA (principal component analysis) and LDA (linear discriminant analysis). The low dimensional PCA and LDA features were then classified by a Bayesian classifiers and 1-nearest neighbor (kNN) classifier. Figure 14 summarize the overall process-flow of the set of experiments.

![Figure 14: Process flow of the experiment](image-url)
Furthermore, the combination of the low dimensional, discriminative PCA and LDA features along with powerful [50] Bayesian classifiers allow us to achieve significant improvement in recognition accuracy. This is because Bayesian classifiers have the flexibility to incorporate prior information, and can predict how a system’s performance will change when going from one environment to another or when going from one type of testing to another [48]. And \( k\)-NN is very effective simple classifier with noise reduction capabilities and instance-based learning [46]. The Bayesian linear and quadratic discriminant classifier uses Bayesian decision rule for classifying a set of learned feature vectors to a class. While the linear classifier fits a multivariate normal density to each group, with a pooled estimate of covariance, the quadratic discriminant classifier fits MVN (multivariate normal) densities with covariance estimates stratified by group. Both methods use likelihood ratios to assign observations to groups. Given a set of classes \( M \) characterized by a set of known parameters in model \( \Omega \) a set of extracted feature vector \( X \) belongs to the class which has the highest probability. This is shown in Eq.(8) and is known as Bayesian decision rule.

\[
X \in M_i \quad P(M_i|X,\Omega) \geq P(M_l|X,\Omega) \quad \forall l \neq k \quad \text{………………………………………………………(8)}
\]

To calculate the a-posterior probability shown, I used Bayesian law of statistics which finally by assuming that features are distributed normally, leads to a quadratic classifier format known as Bayesian Quadratic classifier [51]. The model \( \Omega \) consists of the mean and the covariance of our training vectors, and likelihoods are calculated as stated above.

**3.8.1.1 Experiment with “ear-only”**

For the first set of experiments I applied PCA-LDA transformation and performed classification with Bayesian (linear/quadratic) and \( k\)-nearest neighbor classifiers. Table 8 shows the recognition accuracies.
Table 8: Ear recognition rate by using PCA-LDA

<table>
<thead>
<tr>
<th>No</th>
<th>Classifier (s)</th>
<th>Recognition Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bayesian Linear</td>
<td>94%</td>
</tr>
<tr>
<td>2</td>
<td>Bayesian Quadratic</td>
<td>94%</td>
</tr>
<tr>
<td>3</td>
<td>kNN</td>
<td>88%</td>
</tr>
</tbody>
</table>

From this experiment, I received 94% recognition accuracy for Bayesian-linear classifier, 94% accuracy for Bayesian quadratic and 88% accuracy for 1-NN classifier. Though I expected 100% accuracy, I found that quality of images was very poor, resulting in failure to recognize some poor quality ear image. However, PCA was still able to model the low resolution image pretty good.

![Figure 15: PCA Eigen Value of ear recognition](image)

3.8.1.2 Experiment with “Side-face and Ear”
Here I performed experiments with fusion of face and ear, and recognition accuracies achieved is shown in Table 9. An improvement in recognition performance was achieved with face and ear fusion resulting an accuracy of 100% for all three classifiers. This is basically feature level fusion of ear and face.

Table 9: Fusion result of ear and face

<table>
<thead>
<tr>
<th>No</th>
<th>Classifier (s)</th>
<th>Recognition Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bayesian Linear</td>
<td>100%</td>
</tr>
<tr>
<td>2</td>
<td>Bayesian Quadratic</td>
<td>100%</td>
</tr>
<tr>
<td>3</td>
<td>kNN</td>
<td>100%</td>
</tr>
</tbody>
</table>
This is really a promising result. Once I received 100% recognition rate for all three classifiers, I examined fusion of all three traits. The scatter for face-ear fusion experiment has shown in Figure 16 above.

3.8.1.3 Experiment with “Gait only”
For this experiment, as usual I used all 72 sequences for 18 different people. I extracted gait only (lower part of the body) image from the video sequence. And I applied PCA-LDA with Bayesian linear, Bayesian quadratic and kNN classifier. The accuracy achieved shown in Table 10.

<table>
<thead>
<tr>
<th>No</th>
<th>Classifier (s)</th>
<th>Recognition Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bayesian Linear</td>
<td>95%</td>
</tr>
<tr>
<td>2</td>
<td>Bayesian Quadratic</td>
<td>84%</td>
</tr>
<tr>
<td>3</td>
<td>kNN</td>
<td>89%</td>
</tr>
</tbody>
</table>

The result shows that I received best output by using Bayesian linear classifier which is 95%.

Whereas by using Bayesian quadratic and kNN I achieved 84% and 89% accuracy rate.
respectively. Still, overall I would say, this is good result since this is non-intrusive 
behavioural trait.

![Eigen face of projected gait](image)

**Figure 17: Eigen face of projected gait**

**3.8.1.4 Experiment of Ear, Side-Face and Gait Fusion**

This is our main experiment of this chapter; where I received our final as well as expected 
output. It was feature level fusion of ear-face and gait. Once again, I applied principle
component analysis and linear discriminant analysis with respect to Bayesian classifiers and
kNN classifier. Here I used all 72 sequences for all 18 subjects. And I performed image pre-
processing steps corresponding to background segmentation, cropping, filtering and
histogram equalization of images, same as I did for our previous experiments. The result of
feature level fusion of ear-face-gait shown in table 11 bellow.

**Table 11: Result of feature level fusion of ear, face and gait**

<table>
<thead>
<tr>
<th>No</th>
<th>Classifier(s)</th>
<th>Recognition Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bayesian Linear</td>
<td>89%</td>
</tr>
<tr>
<td>2</td>
<td>Bayesian Quadratic</td>
<td>89%</td>
</tr>
<tr>
<td>3</td>
<td>kNN</td>
<td>84%</td>
</tr>
</tbody>
</table>

Achieved result shows that I received 89% accuracy in recognition by using Bayesian linear
and quadratic classifiers. By using kNN I received a bit poor result then Bayesian classifiers
which is 84%. I achieved 100% correct recognition rate in fusion of ear and side face, but in the fusion of ear, side face and gait seems poor performance in comparing to ear-side face fusion. Well, I realized the gait data provided poor feature in fusion aspect, as this is long range biometric. For my further experiment we will address the issue and I’m confident that I will receive much better result with high quality data. And the collection of quality data is on process. The Figure 18 and 19 shows the canonical correlation and the value of k-mean clustering of the fusion.

Figure 18: CCA (Canonical correlation analysis) Map of Eigen value

Figure 19: k-mean clustering of projected Eigen value
However, the results of face, ear and gait fusion are catastrophic due to poor quality of data. I envisage the approach will work better with use of better quality data. Moreover, our further work involves carrying out experiments with person wearing different clothes and exploring novel methods for identity verification for unconstrained operating environments with less training data. In that I’m expecting to have better result especially in fusion of all three traits. I’m considering some other approaches as well, in regards to new classifier techniques to increase performance and accuracy rate.

3.9 Summary
In this chapter, I have presented a review of current biometric identification technologies and have done number experiment of face and gait biometric traits for next generation biometric technologies. The importance of primary and secondary biometric traits and the role of fusion protocols in addressing the requirements of next generation biometrics are discussed. Then I have investigated novel multimodal identification approaches based on fusion of face and gait biometric cues form low resolution surveillance videos. The proposed approach was based on transforming the features in PCA-LDA subspace, and classification with Multivariate Gaussian (linear and quadratic) classifiers. The experimental evaluation of the proposed scheme on a publicly available database[36] showed that the combined PCA-LDA approach turns out to be a powerful method for capturing the inherent multimodality in walking gait patterns and at the same time discriminating the identity from low resolution video with noisy backgrounds. Further, I extended identification approach based on fusion of ear, face and gait biometric cues form low resolution surveillance videos. The results of face, ear and gait fusion are catastrophic due to poor quality of data. I envisage the approach will work better with use of better quality data. However, our proposed modal on multimodal fusion of face, ear and gait biometric cues for identity verification is suitable for identifying people in low resolution surveillance videos, which a futuristic and next generation solution for real world
identity verification problem. Further, this will allow diffusion of biometric security technologies with better user-acceptability for day to day civilian access control and public surveillance applications. Moreover, next chapter of the thesis involves carrying out experiments with person wearing different clothes and exploring novel methods for identity verification for unconstrained operating environments with less training data. In that I’m expecting to have better result especially in fusion of all three traits. I’m also considering some other approaches as well, in regards to new classifier techniques to enhance the performance, in terms of recognition accuracy, false accept and reject rates.
Chapter 4: Multimodal Face-Gait Fusion with CASIA Gait Database

4.1 Introduction

Human identification from arbitrary views is a very challenging problem, especially when one is walking at a distance. Over the last few years, recognizing identity from gait patterns has become a popular area of research in biometrics and computer vision, and one of the most successful applications of image analysis and understanding. Also, gait recognition is being considered as a next-generation recognition technology, with applicability to many civilian and high security environments such as airports, banks, military bases, car parks, railway stations etc. For these application scenarios, it is not possible to capture the frontal face, and is of low resolution. Hence most of traditional approaches used for face recognition fail; however, several studies have shown that humans can identify a person from a distance from their gait or the way they walk. Even if frontal face is not visible, it is possible to establish the identity of the person using certain static and dynamic cues such as from face, ear, walking style, hand motion during walking etc. If automatic identification systems can be built based on this concept, it will be a great contribution to surveillance and security area.

However, each of these cues or traits captured from long range low resolution surveillance videos on its own are not powerful enough for ascertaining identity. A combination or fusion of each of them, along with an automatic processing technique can result in satisfactory recognition accuracies. With this experiment, I propose usage of full profile silhouettes of persons without frontal faces from visible range and infrared range, for capturing inherent multi-modality available from static and dynamic cues from the gait patterns of the walking human. This also addresses the problems with frontal faces, such as vulnerability to pose, illumination and expression variations. In addition, one of the biggest shortcomings of frontal face is; user cooperation is mandatory upon data collection. On other hand, long range biometric information from surveillance videos captures several biometric traits such as side face, ear, body shape, and gait, which are a combination of physiological and behavioral
biometrics resulting in robust identification approaches. Further, by using certain automatic processing techniques for extracting salient features based on multivariate statistical techniques and learning classifiers, it is possible to enhance the performance in real world operating scenarios. Here, I use simple feature extraction techniques based on principle component analysis (PCA) and linear discriminant analysis (LDA) with different types of learning classifiers. The experimental evaluation of the scheme on a publicly available CASIA [52] database with visible and infra-red gait information shows promising performance improvement.

4.2 Literature Search
For experimental evaluation of our proposed face-gait identification scheme, I used CASIA Gait Database collected by Institute of Automation, Chinese Academy of Sciences [52]. It is a large multi-view gait database, which is created in January 2005. There are more than 300 subjects. I used two different set of data known as dataset B and Dataset C. Dataset B was captured from 11 views with normal video camera, and 11 different views know as view angles. I used the data captured only in 90 degree view angle. The dataset C was captured with an infrared (thermal) camera. It takes into account four walking conditions: normal walking, slow walking, fast walking, and normal walking with a bag. The videos were all captured at night. Figure 20 shows the sample images in different view angles.
However, I used 50 subjects with a set of extracted silhouettes from Dataset B and another set of extracted silhouettes from Dataset C. Each set consists of 16 images and in total 1600 images for 100 subjects (people). Figure 21 shows the extracted silhouettes from datasets B and C.
Further, I extracted the feature vector for each of the dataset separately by using PCA (principal component analysis) and Linear Discriminant Analysis (LDA). And classified with different classifiers. So the tests involved PCA-MLP, LDA-MLP, PCA-SMO, LDA-SMO, LDA-Naïve Bayes, LDA-J48. Each of them is described briefly in next section. I applied all mentioned classifier by using WEKA machine learning software [53]

4.3 Methodology

**PCA-LDA:** Principle component analysis is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analysing data. The other main advantage of PCA is that once I have found these patterns in the data, and I can compress the data, e.g. by reducing the number of dimensions, without much loss of information. Basically this technique used in image compression [54]. In the image analysis it works like;

\[ X = (x_1, x_2, x_3 \ldots N_2) \]

\[ \text{where the rows of pixels in the image are placed one after the other to form a one dimensional image. Each image is } N \text{ pixels high by } N \text{ pixels wide. For each image it creates an image vector. And then it counts all the images together in one big image-matrix like;} \]

\[ \text{Matrix} = (v_1, v_2, v_3 \ldots v_N) \]

\[ \text{(10)} \]

On the other hand, the LDA also closely related to principal component analysis (PCA) and factor analysis in that they both look for linear combinations of variables which best explain the data. LDA explicitly attempts to model the difference between the classes of data. PCA on the other hand does not take into account any difference in class, and factor analysis builds the feature combinations based on differences rather than similarities. Discriminant analysis is also different from factor analysis in that it is not an interdependence technique: a distinction between independent variables and dependent variables (also called criterion
variables) must be made. LDA works when the measurements made on independent variables for each observation are continuous quantities. When dealing with categorical independent variables, the equivalent technique is discriminant correspondence analysis [55]. And in our experiment, LDA shows prominent than PCA. However, The Figure 22 shows the extracted feature (Eigen value) using PCA in Dataset B, and next Section describes several classifiers I examined.

![Figure 22: The extracted feature (Eigen value) using PCA in Dataset B](image)

### 4.3.1 Multi-Layer Perceptron

Multi-Layer perceptron (MLP) is a feed forward neural network with one or more layers between input and output layer. Feed forward means that data flows in one direction from input to output layer (forward). This type of network is trained with the back propagation learning algorithm. MLPs are widely used for pattern classification, recognition, prediction and approximation. Multi-Layer Perceptron can solve problems which are not linearly separable [56]. Figure 23 shows that sample layer representation in MLP.
The network diagram shown above is a full-connected, three layer, feed-forward, perceptron neural network. “Fully connected” means that the output from each input and hidden neuron is distributed to all of the neurons in the following layer. “Feed forward” means that the values only move from input to hidden to output layers; no values are fed back to earlier layers (a Recurrent Network allows values to be fed backward) [57]. In our experiment I had 49 input layer, 800 hidden layer (for each data set) and 50 output layer. This is basically based on dimension, instance and the given class.

4.3.2 J48 Classifier
J48 classifier is same as C4.5 algorithm. It is used to generate a decision tree developed by Ross Quinlan. C4.5 is an extension of Quinlan’s earlier ID3 algorithm. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier. It builds decision trees from a set of training data in the same way as ID3, using the concept of information entropy. The training data is a set $S = s_1, s_2, \ldots$ of already classified samples. Each sample $s_i = x_1, x_2, \ldots$ is a vector where $x_1, x_2, \ldots$ represent attributes or features of the sample. The training data is augmented with a vector $C = c_1, c_2, \ldots$ where $c_1, c_2, \ldots$ represent the class to which each sample belongs [58]. At each node of the tree, C4.5 chooses one attribute of the data that most effectively splits its set of samples into
subsets enriched in one class or the other. Its criterion is the normalized information gain
(difference in entropy) that results from choosing an attribute for splitting the data. The
attribute with the highest normalized information gain is chosen to make the decision. The
C4.5 algorithm then recourses on the smaller sub lists [59].

4.3.3 Naïve Bayes Classifier
A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes’ theorem
with strong (naive) independence assumptions. A more descriptive term for the underlying
probability model would be “independent feature model”. In simple terms, a naive Bayes
classifier assumes that the presence (or absence) of a particular feature of a class is unrelated
to the presence (or absence) of any other feature, given the class variable. In plain English it
works like[58];
Posterior = (Prior*Likelihood)/Evidence ............... (11)
All model parameters (i.e., class priors and feature probability distributions) can be
approximated with relative frequencies from the training set. These are maximum likelihood
estimates of the probabilities. A class prior may be calculated by assuming equiprobable
classes (i.e., priors = 1 / (number of classes)), or by calculating an estimate for the class
probability from the training set (i.e., (prior for a given class) = (number of samples in the
class) / (total number of samples)). To estimate the parameters for a feature’s distribution, one
must assume a distribution or generate nonparametric models for the features from the
training set [58].

4.3.4 SMO Classifier
Finally I examined the SVM classifier with SMO. The Sequential Minimal Optimization
(SMO) is a simple algorithm in the machine learning area. SMO decomposes the overall QP
problem into QP sub-problems, using Osuna’s theorem to ensure convergence [57]. Unlike
the other methods, SMO chooses to solve the smallest possible optimization problem at every
step. The advantage of SMO lies in the fact that solving for multi instance multipliers can be
done analytically. In addition, SMO requires no extra matrix storage at all. There are two components to SMO: an analytic method for solving for the two Lagrange multipliers, and a heuristic for choosing which multipliers to optimize [60].

\[ y_1 \neq y_2 \Rightarrow \alpha_1 - \alpha_2 = k \]  
\[ y_1 = y_2 \Rightarrow \alpha_1 + \alpha_2 = k \]

However, the multi instance multipliers must fulfill all of the constraints of the full problem. The linear equality constraint causes them to lie on a diagonal line. Therefore, one step of SMO must find an optimum of the objective function on a diagonal line segment [60].

![Normalized feature vectors](image)

Figure 24: Normalized feature vectors

Once I normalized the data (As shown Figure 24), I examined different classifiers for their recognition accuracy.

### 4.4 Experiments
For this experiment, I used Dataset C which has been captured on infrared camera.

For feature extraction I used linear discriminant analysis (LDA) technique. The result shows J48 classifier providing poor result in comparison to all other classifiers which are around 64%. Further, MLP and SMO show significant improvement in classification and they classification accuracy is 94% and 93% respectively. Moreover, to compare both dataset B
and C I applied PCA-MLP and LDA-MLP separately. Figure 25 shows the result of Dataset B with LDA-MLP and PCA-MLP.

Figure 25: Accuracy difference in applied classifiers

Figure 26: Accuracy differences in PCA-LDA with MLP
The result (Figure 25 and 26) shows, more than 92% accuracy was achieved by using LDA, whereas, only 79.5% accuracy was achieved by using PCA. LDA features providing good results as compared to PCA features. Figure 26 shows the result achieved from Dataset C Moreover, the results show; that I received more than 94% accuracy with LDA features by using MLP classifier, and 83% accuracy with PCA features. These experiments show that overall LDA features working much better as compared to PCA features, for both Dataset B (Visible) and Dataset C (Infrared). The result shows the dataset taken by infrared camera providing 83% and 94% for PCA, LDA respectively.

Furthermore, to validate our scheme I applied different folds of cross-validation (Figure 27). Cross-validation, sometimes called rotation estimation is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. One round of cross-validation involves partitioning a sample of data into complementary subsets, performing the analysis on one subset (called the training set), and validating the analysis on the other subset (called the validation set or testing set). To reduce variability, multiple rounds of cross validation are performed using different
partitions, and the validation results are averaged over the rounds [57]. In this experiment I
applied 5, 10, 15, 20, and 30 folds cross-validations. I found that the overall accuracy
changing over cross validation size. Figure 28 shows the result achieved from dataset B

![Result on Dataset B with different size of Cross-Validation Fold](image)

*Figure 28: Accuracy difference in different folds LDA-MLP with dataset B*

The result is fluctuating on size of folds. It shows 10 folds provided better result in compare
to other applied folds. The best result is around 93% with 10 folds and 15 folds cross-
validation given poor result that is around 89% accuracy. However, figure 29 shows the result
in dataset C with different size of cross-validation.
With dataset C 15 folds provided better result which is around 94.5% and 5 folds cross-validation provided 93.3% which the lowest accuracy in this experiment. The dimensions of the PCA and LDA features for examining the influence of different folds of cross-validation were 49 dimensions. The final set of experiments involved influence of different dimensions of PCA and LDA features, as dimensionality can affect the speed of the recognition system. For this set of experiments, I fixed number of folds for cross validation to 10 folds, as this seems to be optimal from Figure 28 and 29. The results are shown in Table 12 here for SMO classifier. As can be seen in Table 12, the best recognition accuracy is achieved for dataset C with LDA features with 40 dimensions and is of the order of 93.88%. The LDA features perform better with lesser dimensions as compared to PCA features. To summarize our experimental evolution I can say that; performance depends on different parameters, such as type of features, type of classifiers, dimensionality of the features and number of cross-validation folds used. As can be seen, for dataset C (which has been captured in an infrared camera), the linear discriminant analysis (LDA) - Multilayer Perceptron (MLP) - 15 folds cross-validation turns to be the best combination.
<table>
<thead>
<tr>
<th>Dataset</th>
<th>Features</th>
<th>Dimensions</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset B</td>
<td>PCA</td>
<td>10</td>
<td>4.39</td>
</tr>
<tr>
<td>Dataset B</td>
<td>PCA</td>
<td>20</td>
<td>10.38</td>
</tr>
<tr>
<td>Dataset B</td>
<td>PCA</td>
<td>30</td>
<td>19.88</td>
</tr>
<tr>
<td>Dataset B</td>
<td>PCA</td>
<td>40</td>
<td>37.76</td>
</tr>
<tr>
<td>Dataset B</td>
<td>PCA</td>
<td>49</td>
<td>52.63</td>
</tr>
<tr>
<td>Dataset B</td>
<td>LDA</td>
<td>10</td>
<td>2.13</td>
</tr>
<tr>
<td>Dataset B</td>
<td>LDA</td>
<td>20</td>
<td>17.88</td>
</tr>
<tr>
<td>Dataset B</td>
<td>LDA</td>
<td>30</td>
<td>39.63</td>
</tr>
<tr>
<td>Dataset B</td>
<td>LDA</td>
<td>40</td>
<td>54.66</td>
</tr>
<tr>
<td>Dataset B</td>
<td>LDA</td>
<td>49</td>
<td>68.13</td>
</tr>
<tr>
<td>Dataset C</td>
<td>PCA</td>
<td>10</td>
<td>76</td>
</tr>
<tr>
<td>Dataset C</td>
<td>PCA</td>
<td>20</td>
<td>88</td>
</tr>
<tr>
<td>Dataset C</td>
<td>PCA</td>
<td>30</td>
<td>88</td>
</tr>
<tr>
<td>Dataset C</td>
<td>PCA</td>
<td>40</td>
<td>90.63</td>
</tr>
<tr>
<td>Dataset C</td>
<td>PCA</td>
<td>49</td>
<td>90.68</td>
</tr>
<tr>
<td>Dataset C</td>
<td>PCA</td>
<td>10</td>
<td>91.25</td>
</tr>
<tr>
<td>Dataset C</td>
<td>LDA</td>
<td>20</td>
<td>86.88</td>
</tr>
<tr>
<td>Dataset C</td>
<td>LDA</td>
<td>30</td>
<td>93.23</td>
</tr>
<tr>
<td>Dataset C</td>
<td>LDA</td>
<td>40</td>
<td>93.88</td>
</tr>
<tr>
<td>Dataset B</td>
<td>LDA</td>
<td>49</td>
<td>93.38</td>
</tr>
</tbody>
</table>

4.5 2D Fusion (In View-Point)

For experimental evaluation of our proposed face-gait identification scheme, I used CASIA Gait Database collected by Institute of Automation, Chinese Academy of Sciences [52]. It is a large multi-view gait database, which is created in January 2005. There are more than 300 subjects. I used two different set of data known as dataset B and Dataset C. Dataset B was captured in 90 degree view point and dataset C captured in 126 degree view point. It takes into account four walking conditions: normal walking, slow walking, fast walking, and normal walking with a bag. The videos were all captured at night. Figure 30 shows some of the sample images from the walking video sequences. In total I used 100 subjects with a set of extracted silhouettes from Dataset B and another set of extracted silhouettes from Dataset C. Each subject consists of 16 images and in total 1600 images for 100 subjects (people). Figure 30 shows the extracted silhouettes from datasets B and C.
I extracted the feature vectors from each of the dataset images separately by using Linear Discriminant Analysis (LDA). The LDA is a multivariate statistical analysis technique similar to principal component analysis (PCA) and factor analysis - they both look for linear combinations of variables which best explain the data. LDA explicitly attempts to model the difference between the classes of data. PCA on the other hand does not take into account any difference in class, and factor analysis builds the feature combinations based on differences rather than similarities. Discriminant analysis is also different from factor analysis in that it is not an interdependence technique: a distinction between independent variables and dependent variables (also called criterion variables) must be made. LDA works when the measurements made on independent variables for each observation are continuous quantities. When dealing with categorical independent variables, the equivalent technique is discriminant correspondence analysis [55]. Both LDA and PCA are dimensionality reduction techniques and allow the variations to be modelled with few discriminant or principal vectors. Figure 31 shows how the first few features model significant variations for Dataset B images.
Once I extracted the feature vectors, I fused the features from different view point data sets. For experimental validation, I divided the data sets into training, validation and test data sets. During the training phase, the template for each person was constructed using single mode and fused LDA features from each view point data set. For classification, I used “logistic function” implemented in WEKA machine learning software [53]. The logistic function or logistic curve is a common sigmoid curve, given its name in 1844 or 1845 by Pierre François Verhulst who studied it in relation to population growth. It can model the "S-shaped" curve (abbreviated S-curve) of growth of some population $P$. The initial stage of growth is approximately exponential; then, as saturation begins, the growth slows, and at maturity, growth stops. Sometime the Logistic functions combine, in one neat package, two characteristic kinds of exponential growth. Figure 32 shows the combination of logistic function.
The first kind of exponential growth is the familiar pattern of increase at an increasing rate. Since the growth is exponential, the growth rate is actually proportional to the size of the function's value. The second kind of exponential growth is usually called bounded exponential growth. It's really just a clever trick: It takes a decaying exponential and subtracts it from a fixed bound. As the decaying exponential dies out, the difference rises up to the bound. This kind of function models growth that is limited by some fixed capacity [61]. Logistic functions combine the first kind of exponential growth, when the outputs are small, with the second kind of exponential growth, when the outputs near capacity. In our experiments the logistic function works as 2nd kind of exponential as I was looking on nearest capacity. I also examined different folds of cross-validation. Cross-validation, sometimes called rotation estimation is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. One round of cross-validation involves partitioning a sample of data into complementary subsets, performing the analysis on one subset (called the training set), and validating the analysis on the other subset (called the validation set or testing set). To reduce variability, multiple rounds of cross-validation are performed using different partitions, and
the validation results are averaged over the rounds [62]. In total I have three different experimental set ups namely: identification in single mode 90 degree view point, identification in single mode 126 degree view point and identification in fusion. Figure 33 shows the identification in 90 degree view point.

The figure shows I received around 84.5% correct identification by using LDA-Logistic approach from the data has captured in 90 degree view point. And only 15.5% has been identified with wrong identification. On the other hand, the data captured in 126 degree view point providing us better result than the 90 degree view point. Figure 34 shows the result achieved in 126 degree view point.
The result shows, I received significant improvement in recognition accuracy for the data, which has been captured in 126 degree view point. By using dataset C (126 degree), I received almost 88.88% correct identification for a large data set (over 50 subjects or persons). And wrong/incorrect identification rate is around 11.12% which is still large enough in real world scenario. Finally I moved to our main experiment that is feature level fusion of both datasets. Figure 35 represents the identification in the fusion.

As can be seen in Figure 35, a significant improvement in recognition accuracy is achieved with view point fusion. Whereas for dataset B, recognition accuracy was 84%, and for dataset C it was 88% identification, the fusion of two viewpoints results in 91% accuracy. This is a significant improvement. Further work involves development of new feature extraction and classifier techniques for enhancing the recognition accuracy.

4.6 3D Fusion (In View-Point)
For experimental evaluation of the proposed multiview gait fusion schemes, I used CASIA Gait Database collected by Institute of Automation, Chinese Academy of Sciences [52]. It is a large multi-view gait database, which is created in January 2005. There are more than 300 subjects. I used three (3) different datasets known as dataset A (36 degree view point) dataset
B (90 degree view point) and Dataset C (126 degree view point). All data was captured with normal video camera in 11 different views known as view angles. It takes into account four walking conditions: normal walking, slow walking, fast walking, and normal walking with a bag. All of my data here in this experiment taken from normal walking with free hand. The videos were all captured at night. For all the experiments, I used 50 subjects from each of the dataset. It means, I used 50 subjects of extracted silhouettes from Dataset A, 50 subjects from B and 50 subjects from C. Each subject consists of 16 images and in total 2400 images for 150 subjects.

For each of the images in these data sets, I extracted the feature vectors in lower dimensional subspaces separately by using PCA (principal component analysis) and Linear Discriminant Analysis (LDA), and used a learning classifier based on well know multi-layer perceptron (MLP) for classifying each person ID. Our multiview fusion experiments involved identity recognition in LDA-MLP subspace for dataset (unimodal) and fusion of multiple views. The details of LDA subspace for extracting discriminating features is described next.

4.6.1 Linear Discriminant Analysis
The Linear Discriminant Analysis (LDA) similar to principal component analysis (PCA) and factor analysis, looks for linear combinations of variables which can best explain the data. LDA explicitly attempts to model the difference between the classes of data. PCA on the other hand does not take into account any difference in class, and factor analysis builds the feature combinations based on differences rather than similarities. Discriminant analysis is also different from factor analysis in that it is not an interdependence technique: a distinction between independent variables and dependent variables (also called criterion variables) must be made. LDA works when the measurements made on independent variables for each observation are continuous quantities. When dealing with categorical independent variables, the equivalent technique is discriminant correspondence analysis [55]. In our experiment
LDA shows very promising as LDA model the difference between class and data. Figure 36 shows the extracted feature using LDA.

![Figure 36: LDA values extracted from silhouettes](image)

4.6.2 Multi-Layer Perceptron
Multi Layer perceptron (MLP) is a feed forward neural network with one or more layers between input and output layer. Feed forward means that data flows in one direction from input to output layer (forward). This type of network is trained with the back-propagation learning algorithm. MLPs are widely used for pattern classification, recognition, prediction and approximation. Multi Layer Perceptron can solve problems which are not linearly separable [57]. In our experiments I had 49 input layer, 800 hidden layer (for each data set) and 50 output layer. This is basically based on dimensions, instances and the classes of the dataset. Figure 37 shows the network architecture with MLP, where the green baton to very left; represents dimensions of LDA feature vector, the yellow baton in very right; represents each class (person). The details of the experiments is described in the next section.
4.6.3 Experimental Results and Discussion

The experiments involved a training phase and a test phase. I used a 10-fold cross-validation for dividing the complete data from into training and test subsets. With 10-fold cross-validation, the original dataset is randomly partitioned into 10 subsets. Of the k subsets, a single subset was retained as the validation data for testing the model, and the remaining 9 subsets were used as training data. The cross-validation process is then repeated 10 times (the folds), with each of the 10 subsets used exactly once as the validation data. The 10 results from the folds were then averaged to produce a single estimation. I found that advantage of this method over repeated random sub-sampling is that all observations could be used for both training and validation/testing, and each observation could be used for validation exactly once. In training phase, I built the gait templates for each person using LDA feature vectors for each of the dataset images (Datasets A, B and C) and trained the MLP classifier. In test phase the LDA feature vectors from unseen images in training set were classified with MLP classifier for each of the datasets separately (datasets A, B, C) and by fusion of multiple views. Figure 38 shows the rate of identification in 36 degree view point.
The figure shows high level of accuracy with the proposed scheme, for data captured in 36 degree view point. I achieved 98% correct identification by using LDA-MLP approach. And only 2% has been identified with wrong/incorrect identification. On the other hand, the data captured in 90 degree view point resulted in poor results as compare to the data from 36 degree view point. This could be due to difficulty in capturing the identity specific information from 90 degree view point as compared to 36 degrees. Figure 39 shows the results achieved with the data from 90 degree view point.
The result shows, I received 84.5% correct identification for a large data set which has captured in 90 degree view point. And wrong/incorrect identification rate is around 14.5% which is quite large for real world scenario. Figure 40 represents the identification for dataset C (126 degree view point). It can be seen from this figure that it was possible to achieve 88.88% correct identification with the data captured in 123 degree view point. And 11.12% identified were wrongly identified.
After three (3) successful single mode experiments I combined data from all views. I performed feature level fusion of all three extracted set of LDA features, and Figure 41 shows the results of multi-view feature fusion based on gait images from surveillance videos.

![Multi-view Fusion](image)

*Figure 41: Result of 3-D fusion*

As can be seen from Figure 41, feature level fusion of multiple views results in a significant improvement in correct identification rate as compared to single views, with 99% accuracy for fusion of multiple views. Further, accuracy of each class individually was also good with excellent true positive (TP) rates. The figures for detailed accuracy are shown in the Figure 42. As mentioned earlier, each class in Figure 42 represents each individual. Finally, to summarize our experimental validation I can say that; by using multiple views of surveillance video footage with long range videos (without detailed face images), it is possible to perform large scale identification with high level of accuracy, using simple subspace features (LDA) and classifier techniques (MLP). Such simple approaches can lead to real time and real world intelligent video surveillance systems - the beginning of a new dimension of security systems in public surveillance. Our experimental efforts reported here shows the importance of multiview images from several cameras and feature level fusion of multiple views as an efficient gait biometric identification.
### Detailed Accuracy By Class

<table>
<thead>
<tr>
<th>Class</th>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>ROC Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>s1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s4</td>
<td>0.875</td>
<td>0.003</td>
<td>0.875</td>
<td>0.875</td>
<td>0.875</td>
<td>0.994</td>
</tr>
<tr>
<td>s5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s6</td>
<td>0.875</td>
<td>0.003</td>
<td>0.875</td>
<td>0.875</td>
<td>0.875</td>
<td>0.998</td>
</tr>
<tr>
<td>s7</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s8</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s9</td>
<td>0.938</td>
<td>0</td>
<td>1</td>
<td>0.938</td>
<td>0.966</td>
<td>1</td>
</tr>
<tr>
<td>s10</td>
<td>0.938</td>
<td>0.001</td>
<td>0.938</td>
<td>0.938</td>
<td>0.938</td>
<td>1</td>
</tr>
<tr>
<td>s11</td>
<td>0.938</td>
<td>0</td>
<td>1</td>
<td>0.938</td>
<td>0.966</td>
<td>1</td>
</tr>
<tr>
<td>s12</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s13</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s14</td>
<td>1</td>
<td>0.001</td>
<td>0.941</td>
<td>1</td>
<td>0.97</td>
<td>1</td>
</tr>
<tr>
<td>s15</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s16</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s17</td>
<td>1</td>
<td>0.001</td>
<td>0.941</td>
<td>1</td>
<td>0.97</td>
<td>1</td>
</tr>
<tr>
<td>s18</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s19</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s20</td>
<td>0.938</td>
<td>0</td>
<td>1</td>
<td>0.938</td>
<td>0.966</td>
<td>1</td>
</tr>
<tr>
<td>s21</td>
<td>1</td>
<td>0.003</td>
<td>0.889</td>
<td>1</td>
<td>0.941</td>
<td>1</td>
</tr>
<tr>
<td>s22</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s23</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s24</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s25</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s26</td>
<td>1</td>
<td>0.001</td>
<td>0.941</td>
<td>1</td>
<td>0.97</td>
<td>1</td>
</tr>
<tr>
<td>s27</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s28</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s29</td>
<td>0.938</td>
<td>0</td>
<td>1</td>
<td>0.938</td>
<td>0.966</td>
<td>1</td>
</tr>
<tr>
<td>s30</td>
<td>1</td>
<td>0.003</td>
<td>0.889</td>
<td>1</td>
<td>0.941</td>
<td>1</td>
</tr>
<tr>
<td>s31</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s32</td>
<td>0.875</td>
<td>0</td>
<td>1</td>
<td>0.875</td>
<td>0.933</td>
<td>1</td>
</tr>
<tr>
<td>s33</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s34</td>
<td>0.941</td>
<td>0</td>
<td>1</td>
<td>0.941</td>
<td>0.97</td>
<td>0.951</td>
</tr>
<tr>
<td>s35</td>
<td>1</td>
<td>0.001</td>
<td>0.938</td>
<td>1</td>
<td>0.969</td>
<td>0.999</td>
</tr>
<tr>
<td>s36</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s37</td>
<td>0.938</td>
<td>0</td>
<td>1</td>
<td>0.938</td>
<td>0.966</td>
<td>1</td>
</tr>
<tr>
<td>s38</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s39</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s40</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s41</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s42</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s43</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s44</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s45</td>
<td>0.938</td>
<td>0</td>
<td>1</td>
<td>0.938</td>
<td>0.966</td>
<td>1</td>
</tr>
<tr>
<td>s46</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s47</td>
<td>0.938</td>
<td>0.004</td>
<td>0.833</td>
<td>0.938</td>
<td>0.882</td>
<td>0.999</td>
</tr>
<tr>
<td>s48</td>
<td>1</td>
<td>0.001</td>
<td>0.941</td>
<td>1</td>
<td>0.97</td>
<td>1</td>
</tr>
<tr>
<td>s49</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>s50</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
4.7 3D Fusion (In Walking Pattern /Style)
For experimental evaluation of the proposed multi-mode gait fusion scheme, I used CASIA Gait Database collected by Institute of Automation, Chinese Academy of Sciences [52]. It is a large multi-view gait database, which is created in January 2005. There are more than 300 subjects. I used three (3) different datasets known as dataset A (Normal walking) dataset B (Fast walking) and Dataset C (Walking with bag). All data was captured with infrared camera. The videos were all captured at night.

For all the experiments, I used 50 subjects from each of the dataset. It means, I used 50 subjects of extracted silhouettes from Dataset A, 50 subjects from B and 50 subjects from C. Each subject consists of 16 images and in total 2400 images for 150 subjects. However, Figure 43 represents the experimental process flow.

![Image Flow Chart](image)

Further, for each of the images in these data sets, I extracted the feature vectors in lower dimensional subspaces by using Linear Discriminant Analysis (LDA), and used a well known learning classifier “Bagging” for classifying each person ID. Our multi-mode walking fusion experiments involved identity recognition in LDA-Bagging subspace for dataset (unimodal) and fusion of multiple mode. The details of LDA subspace for extracting discriminating features is described next.
4.7.1 Linear Discriminant Analysis
The Linear Discriminant Analysis (LDA) similar to principal component analysis (PCA) and factor analysis looks for linear combinations of variables which can best explain the data. LDA explicitly attempts to model the difference between the classes of data. PCA on the other hand does not take into account any difference in class, and factor analysis builds the feature combinations based on differences rather than similarities. Discriminant analysis is also different from factor analysis in that it is not an interdependence technique: a distinction between independent variables and dependent variables (also called criterion variables) must be made. LDA works when the measurements made on independent variables for each observation are continuous quantities. When dealing with categorical independent variables, the equivalent technique is discriminant correspondence analysis [55]. In our experiment LDA shows very promising as LDA model the difference between class and data. Figures 44 and 45 shows the extracted Eigen value and Eigen faces using LDA

Figure 44: LDA Eigen values extracted from silhouettes
4.7.2 Learning Classifier “Bagging”
From the extracted Eigen value to identify a person I used a learning classifier called bagging. The bagging classifier is a "bootstrap" [63] and ensemble method that creates individuals for its ensemble by training each classifier on a random redistribution of the training set [64]. Research shows that Bagging is effective on "unstable" learning algorithms where small changes in the training set result in large changes in predictions [63]. It also reduces variance and helps to avoid over fitting, and that is the main significance of the classifier to implement into our project. Basically it works like kNN classifier, but it is superior in higher dimensions. Bagging in pattern recognition is not new, but the combination of LDA-Bagging provided promising output in this experiment.

4.7.3 Experimental Results and Discussion
The experiments involved a training phase and a test phase. I used a 10-fold cross-validation for dividing the complete data from into training and test subsets. With 10-fold cross-validation, the original dataset is randomly partitioned into 10 subsets. Of the k subsets, a single subset was retained as the validation data for testing the model, and the remaining 9 subsets were used as training data. The cross-validation process is then repeated 10 times (the folds), with each of the 10 subsets used exactly once as the validation data. The 10 results from the folds were then averaged to produce a single estimation. I found that advantage of this method over repeated random sub-sampling is that all observations could be used for
training and validation/testing, and each observation could be used for validation exactly once. In training phase, I built the gait templates for each person using LDA feature vectors for each of the dataset images (Datasets A, B and C) and trained the bagging classifier. In test phase the LDA feature vectors from unseen images in training set were classified with bagging classifier for each of the datasets separately (datasets A, B, C) and by fusion of multiple-mode of walking. Table 13 shows the rate of true identification in the set of experiment

<table>
<thead>
<tr>
<th>No</th>
<th>Algorithm</th>
<th>Dataset</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LDA-Bagging</td>
<td>Normal Walking</td>
<td>82%</td>
</tr>
<tr>
<td>3</td>
<td>LDA-Bagging</td>
<td>Fast Walking</td>
<td>84.88%</td>
</tr>
<tr>
<td>4</td>
<td>LDA-Bagging</td>
<td>Walking with Bag</td>
<td>88.12%</td>
</tr>
<tr>
<td>5</td>
<td>LDA-Bagging</td>
<td>Fusion of Normal Walking and Fast Walking</td>
<td>91.38%</td>
</tr>
<tr>
<td>6</td>
<td>LDA-Bagging</td>
<td>Fusion of Normal Walking and Walking with Bag</td>
<td>90.25%</td>
</tr>
</tbody>
</table>

The result shows high level of accuracy with the proposed scheme, for dataset of “walking with bag”. I achieved 88.12% correct identification by using LDA-Bagging approach. And only 11.88% has been identified with wrong/incorrect identification. On the other hand, the dataset of “fast walking and normal walking” resulted with 85% and 82% respectively. This is quite interesting to see difference in between the mentioned dataset. In fact, now I’m very much clear on shortcomings of identification using gait profile. The result may vary on walking condition. But, most important thing is; it is still identifiable in any walking style. However, after individual testing of each dataset, I moved to fusion of “normal walking and fast walking” and fusion of “normal walking and walking with school bag”. The fusion result shows, fusion of normal walking and fast walking giving us better result in compare to fusion of “normal walking” and “walking with bag”. Still the difference is not much (91.38% and 90.25%). Figure 46 shows the graphical presentation of achieved result.
Moreover, the fusion over individual dataset seems promising, and identification accuracy of each class individually was also good with excellent true positive (TP) rates. The Figure 47 shows the comparison of first 20 class of our experiment.

As I have written early in this chapter, in total I had 50 subjects (person) with 3 different dataset (50*3). Each subject represents a class in our experiment. The figure above clearly
indicating that the true positive (TP) rates are superior over false positive (FP) rate. FP rate seems in the ground in compare to TP. However, in summary of the experiment we can say, by using multi-mode of walking in surveillance video footage with long range videos (without detailed face images), it is possible to perform large scale identification with high level of accuracy, using simple subspace features (LDA) and classifier techniques (bagging). Such simple approaches can lead to real time and real world intelligent video surveillance systems - the beginning of a new dimension of security systems in public surveillance. Our small experimental efforts reported here shows the importance of fusion. It also providing the variation on recognition with different walking style, I believe it will help us to create a benchmark of “next generation surveillance security protocol”.

4.8 Multi camera Fusion (Visible and Infrared Camera)
In this experiment propose a novel human-identification scheme from long range gait profiles in surveillance videos. I investigated the role of multi view gait images acquired from multiple cameras, the importance of infrared and visible range images in ascertaining identity, the impact of multimodal fusion, efficient subspace features and classifier methods, and the role of soft/secondary biometric (walking style) in enhancing the accuracy and robustness of the identification systems, Experimental evaluation of several subspace based gait feature extraction approaches (PCA/LDA) and learning classifier methods (NB/MLP/SVM/SMO) on different datasets from a publicly available gait database CASIA [52], show significant improvement in recognition accuracies with multimodal fusion of multi-view gait images from visible and infrared cameras acquired from video surveillance scenarios.

4.8.1 Multimodal Identification Scheme
For experimental evaluation of our proposed multimodal gait identification scheme, I used CASIA Gait Database collected by Institute of Automation, Chinese Academy of Sciences
It is a large multi-view gait database, which is created in January 2005. There are more than 300 subjects. I used two different set of data known as dataset B and Dataset C. Dataset B was captured from 11 views with normal video camera, and 11 different views known as view angles. I used the data captured only in 90 degree view angle. The dataset C was captured with an infrared (thermal) camera. It takes into account four walking conditions: normal walking, slow walking, fast walking, and normal walking with a bag. The videos were all captured at night. However, I used 50 subjects with a set of extracted silhouettes from Dataset B and another set of extracted silhouettes from Dataset C. Each subject consists of 16 images and in total 1600 images for 100 subjects (people). Figure 48 shows the extracted silhouettes from datasets B and C.

![Extracted silhouettes](image)

I extracted the reduced dimensionality feature vector for each of the dataset separately by suing PCA (principal component analysis) and Linear Discriminant Analysis (LDA), and then have classified with different learning classifiers. Therefore our (cross camera feature
level fusion) experiments involved evaluation of different feature extraction and learning classifier combinations including PCA-MLP, LDA-MLP, PCA-SMO, and LDA-SMO.

4.8.2 Feature Extraction Using PCA- LDA Approach
Principle component analysis is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analysing data. The other main advantage of PCA is that once I have found these patterns in the data, and I can compress the data, e.g. by reducing the number of dimensions, without much loss of information. Basically this technique used in image compression[54]. In the image analysis it works like;

\[ X=(x_1, x_2, x_3 \ldots N_2) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ll
experiment, LDA shows prominent than PCA. Next Section describes several classifiers I examined.

4.8.3 Naive Bayes and MLP Neural Network Classifier

Naive Bayes classifier can serve as a baseline classifier due to its simple probabilistic nature based on applying Bayes' theorem with strong (naive) independence assumptions. In other words, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature, given the class variable. Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without using any Bayesian methods [58]. In spite of their naive design and apparently over-simplified assumptions, naive Bayes classifiers have worked quite well in many complex real-world situations. Multi Layer perceptron (MLP) is a feedforward neural network with one or more layers between input and output layer. Feedforward implies that the data flows in one direction from input to output layer (forward). This type of network is trained with the backpropagation learning algorithm. MLPs are widely used for pattern classification, recognition, prediction and approximation. Multi Layer Perceptron can solve problems which are not linearly separable [57].

4.8.4 SVM and SMO Classifiers

Support Vector Machine (SVM) classifiers perform classification tasks by constructing hyperplanes in a multidimensional space that separates cases of different class labels. A support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. Whereas the original problem may
be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, in SVM, the original finite-dimensional space is mapped into a much higher-dimensional space, presumably making the separation easier in that space. To keep the computational load reasonable, the mappings used by SVM schemes are designed to ensure that dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function selected to suit the problem [65]. The hyperplanes in the higher-dimensional space are defined as the set of points whose inner product with a vector in that space is constant. SMO, on the other hand is an SVM classifier with learning based on Sequential Minimal Optimization (SMO). SMO decomposes the overall QP problem into QP sub-problems, using Osuna’s theorem to ensure convergence [60]. Unlike the other methods, SMO chooses to solve the smallest possible optimization problem at every step. The advantage of SMO lies in the fact that solving for multi instance multipliers can be done analytically. In addition, SMO requires no extra matrix storage at all. There are two components to SMO: an analytic method for solving for the two Lagrange multipliers, and a heuristic for choosing which multipliers to optimize [60].

\[ y_1 \neq y_2 \Rightarrow \alpha_1 - \alpha_2 = k \] .................................................(16)

\[ y_1 = y_2 \Rightarrow \alpha_1 + \alpha_2 = k \] .................................................(17)

However, the multi instance multipliers must fulfil all of the constraints of the full problem. The linear equality constraint causes them to lie on a diagonal line. Therefore, one step of SMO must find an optimum of the objective function on a diagonal line segment [60].

4.8.5 Experiments and Results
Different sets of experiments were performed on two datasets in CASIA database- Dataset B containing visible normal images of walking humans, and Dataset C consisting of infrared images. By using PCA and LDA techniques, I extracted the feature vector for both datasets, training different learning classifiers and performed identification experiments with multiple
fold cross validation in single mode and multimodal fusion mode. I used different combinations of features (for example PCA-Dataset B, PCA-Dataset C, LDA-Dataset B, LDA-Dataset C and the feature level fusion of visible and infrared gait images from Dataset B and Dataset C. Table 14 to Table 18 shows the recognition performance for each set of experiments in terms of recognition accuracy and several statistically significant performance measures such as true positive rate (TPR), false positive rate (FPR), precision, recall and F-measure. All experiments involved either 5 or 10 fold cross validation. Cross-validation is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. One fold of cross-validation involves partitioning a sample of data into complementary subsets (training and testing subsets), performing the analysis on one subset (called the training set), and validating the analysis on the other subset (called the validation set or testing set). To reduce variability, multiple folds of cross-validation are performed using different partitions, and the validation results are averaged over the folds. I examined 5 fold and 10 fold cross-validation for each set of experiments.

Table 14: Classifier performance for visible range dataset (Dataset B) with PCA feature with 50 dimensions (NB – naïve Bayes; MLP – Multilayer Perceptron; TRP – True Positive Rate and FPR – False Positive Rate)

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Folds</th>
<th>Accuracy</th>
<th>TPR</th>
<th>FPR</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB</td>
<td>10</td>
<td>48.63</td>
<td>0.49</td>
<td>0.01</td>
<td>0.49</td>
<td>0.49</td>
<td>0.48</td>
</tr>
<tr>
<td>NB</td>
<td>5</td>
<td>47.68</td>
<td>0.48</td>
<td>0.01</td>
<td>0.49</td>
<td>0.48</td>
<td>0.48</td>
</tr>
<tr>
<td>MLP</td>
<td>10</td>
<td>79.5</td>
<td>0.8</td>
<td>0</td>
<td>0.8</td>
<td>0.8</td>
<td>0.79</td>
</tr>
<tr>
<td>MLP</td>
<td>5</td>
<td>75.13</td>
<td>0.75</td>
<td>0.01</td>
<td>0.76</td>
<td>0.75</td>
<td>0.75</td>
</tr>
</tbody>
</table>

The first set of experiments involve Dataset B (visible range dataset) with 50 dimensional PCA features. As can be seen in Table 14, the recognition accuracy for naïve Bayes classifier with different 10-fold and 5 fold cross-validation is low, with 48.63 % for 10 folds and 47.68 for 5 folds. Using MLP neural net classifier (with backpropagation learning) results in better accuracy with 79.5% for 10 folds and 75.13% for 5 folds. However, the MLP classifier is computational intensive with long train and test times. This could be due to inability of PCA features to discriminate multiple classes (50 classes here) with the available data size or the
structure of the neural network used. The second set of experiments involved use of linear discriminant analysis features and use of support vector machine classifier. As can be seen in Table 15, the naïve Bayes classifier with 50 dimensional LDA features results in significant improvement in performance with 92.5% recognition accuracy as compared to 48.6 % with PCA features for 10 fold cross-validation (CV). With 5 fold CV, the LDA features result in an accuracy of 92.25% as compared to 47.68% for PCA features. Due to computational intensive nature of neural net classifiers, I examined SVM classifier.

For this set of experiments, as SVMs are known to have better generalization ability, are less computation intensive, and are based on sound theory, unlike neural networks whose development has followed a more heuristic path. Other advantages of SVM over neural networks are - whilst ANNs can suffer from multiple local minima, the solution to an SVM is global and unique, and SVMs have a simple geometric interpretation and give a sparse solution. Unlike ANNs, the computational complexity of SVMs does not depend on the dimensionality of the input space. ANNs use empirical risk minimization, whilst SVMs use structural risk minimization. SVMs outperform ANNs often, as they are less prone to over-fitting [60]. However, the performance depends on the kernel used and other SVM parameters. As can be in Table 15, different types of kernels - linear kernel (SVM-L), radial basis function kernel (SVM-RBF), polynomial kernel (SVM-poly) and sigmoidal kernel (SVMsigmoid), result in different recognition accuracies. The SVM with linear kernel

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Folds</th>
<th>Accuracy</th>
<th>TP</th>
<th>FP</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB</td>
<td>10</td>
<td>92.5%</td>
<td>0.93</td>
<td>0</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>NB</td>
<td>5</td>
<td>92.25%</td>
<td>0.92</td>
<td>0</td>
<td>0.93</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>SVM-L</td>
<td>5</td>
<td>81.13%</td>
<td>0.81</td>
<td>0</td>
<td>0.82</td>
<td>0.81</td>
<td>0.81</td>
</tr>
<tr>
<td>SVM-L</td>
<td>10</td>
<td>78.75%</td>
<td>0.78</td>
<td>0</td>
<td>0.81</td>
<td>0.79</td>
<td>0.78</td>
</tr>
<tr>
<td>SVM-RBF</td>
<td>5</td>
<td>31.30%</td>
<td>0.3</td>
<td>0.02</td>
<td>0.74</td>
<td>0.3</td>
<td>0.39</td>
</tr>
<tr>
<td>SVM-Poly</td>
<td>5</td>
<td>27.63%</td>
<td>0.28</td>
<td>0.02</td>
<td>0.75</td>
<td>0.28</td>
<td>0.36</td>
</tr>
<tr>
<td>SVM-Sigmoid</td>
<td>5</td>
<td>29.13%</td>
<td>0.29</td>
<td>0.02</td>
<td>0.74</td>
<td>0.29</td>
<td>0.38</td>
</tr>
</tbody>
</table>
performs best with 81.3% recognition accuracy for 5 fold CV, and has a 78.75% for 10 fold CV. Also, for both naïve Bayes and SVM classifier with linear kernel, the performance with 5 fold cross-validation partition was almost similar to 10 fold cross validation. Hence, for rest of the experiments, I used 5 fold CV partition.

### Table 16: Classifier performance for Infrared range dataset (Dataset C) with LDA feature with 50 dimensions (SVM-L Support Vector Machine-Linear Kernel; RBF- Radial Basis Function Kernel; Poly – Polynomial Kernel; SMO – Sequential Minimal Optimization)

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Feature</th>
<th>Dim</th>
<th>Accuracy</th>
<th>TPR</th>
<th>FPR</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB</td>
<td>PCA</td>
<td>50</td>
<td>56.63%</td>
<td>0.57</td>
<td>0.01</td>
<td>0.59</td>
<td>0.57</td>
<td>0.57</td>
</tr>
<tr>
<td>SVM-L</td>
<td>PCA</td>
<td>50</td>
<td>79.88%</td>
<td>0.8</td>
<td>0</td>
<td>0.81</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td>SVM-L</td>
<td>LDA</td>
<td>50</td>
<td>86.25%</td>
<td>0.86</td>
<td>0</td>
<td>0.88</td>
<td>0.86</td>
<td>0.87</td>
</tr>
<tr>
<td>NB</td>
<td>LDA</td>
<td>50</td>
<td>93.75%</td>
<td>0.94</td>
<td>0</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>NB</td>
<td>LDA</td>
<td>25</td>
<td>93.5%</td>
<td>0.94</td>
<td>0</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>SVM-L</td>
<td>LDA</td>
<td>25</td>
<td>83.25%</td>
<td>0.85</td>
<td>0</td>
<td>0.85</td>
<td>0.83</td>
<td>0.84</td>
</tr>
<tr>
<td>SMO-Poly</td>
<td>LDA</td>
<td>25</td>
<td>94%</td>
<td>0.95</td>
<td>0</td>
<td>0.95</td>
<td>0.94</td>
<td>0.94</td>
</tr>
</tbody>
</table>

For the third set of experiments, I examined Dataset C, the infrared camera gait image dataset, with 5 fold cross validation. As can be seen in Table 16. infrared image dataset performs better than visible range dataset for both PCA and LDA features. The recognition accuracy achieved with 50 dimensional PCA features results is 56.3% for naïve Bayes classifier for Dataset C as compared to 47.68% for Dataset B (Table 14). A similar improvement in performance was achieved with 50-dimensional LDA features resulting in a recognition accuracy of 93.75% for Dataset C as compared to 92.25% for Dataset B. Further, I also examined reduced dimensional LDA features, as LDA features seem to model the identities better, even with large number of classes (50 classes/subjects). As can be seen in Table 16, there is no significant loss of accuracy with reduced dimensional feature vectors. With 25 dimensional LDA feature vector, the recognition accuracy achieved was 93.5% for naïve Bayes classifier (as compared to 93.75% for 50 dimensions) and the accuracies were 83.25% for SVM with linear kernel (86.25%). This has a significant advantage as the reduced dimensional feature vector results in improvement in computational speed. In addition, for
this set of experiments, I examined a different version of SVM classifier – SMO, the SVM with Sequential minimal optimization (SMO). SMO classifier uses an efficient algorithm for solving the optimization problem needed for training of support vector machines, and is known to result in a better performance than a traditional SVM which uses much more complex quadratic optimization problem during training. As can be seen in Table 16, the recognition accuracy achieved with SMO classifier with polynomial kernel is 94% as compared to 93.25% achieved with SVM classifier with linear kernel.

The fourth set of experiments involved the feature level fusion of visible and infrared images from Dataset B and Dataset C. As I found the LDA features to be more discriminatory as compared to PCA, I used LDA features for all fusion experiments. As can be seen in Table 17, the fusion of normal visible camera and infrared camera images is synergistic, resulting in improvement in recognition performance as compared to single mode images. For naïve Bayes classifier, 50- dimensional LDA features result in 98.38% accuracy and 25- dimensional LDA features result in 98.5%. The recognition accuracy achieved with SVM-L (linear kernel) for 50-dim LDA features is 74.88% and 72.5% for 25-dim LDA vector. The SMO version of SVM classifier with polynomial kernel results in 98.25% accuracy for 50-dim LDA vector, and for 25 dimensional LDA features, the accuracy is 97.75%. Once again for fusion mode, SMO with polynomial kernel performs better than traditional SVM with linear kernel. An interesting observation was that the multimodal fusion (feature level) performs a more dominant role as compared to the type of classifier or the type of features, as

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Dim</th>
<th>Accuracy</th>
<th>TPR</th>
<th>FPR</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB</td>
<td>50</td>
<td>98.38%</td>
<td>0.98</td>
<td>0</td>
<td>0.99</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>SVM-L</td>
<td>50</td>
<td>74.88%</td>
<td>0.79</td>
<td>0.01</td>
<td>0.75</td>
<td>0.75</td>
<td>0.97</td>
</tr>
<tr>
<td>SMO-Poly</td>
<td>50</td>
<td>98.25%</td>
<td>0.98</td>
<td>0</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>NB</td>
<td>25</td>
<td>98.50%</td>
<td>0.99</td>
<td>0</td>
<td>0.99</td>
<td>0.99</td>
<td>1</td>
</tr>
<tr>
<td>SVM-L</td>
<td>25</td>
<td>72.50%</td>
<td>0.73</td>
<td>0.01</td>
<td>0.77</td>
<td>0.73</td>
<td>0.73</td>
</tr>
<tr>
<td>SMO-Poly</td>
<td>25</td>
<td>97.75%</td>
<td>0.98</td>
<td>0</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
</tbody>
</table>
irrespective of classifier used (naïve Bayes or SVM), the recognition accuracy is significantly higher with multimodal fusion (higher than 95%).

The final set of experiments involved investigating the role of soft or secondary biometric information, in terms of walking style (fast walking and normal walking) for enhancing the recognition accuracy. The walking style data was available for visible camera images only for all 50 subjects (persons). I used the data for each person; walking in two (2) different styles - fast and normal walking. In this final set of experiments, I examined three different approaches. First, I applied LDA-MLP separately to (1) normal walking data, (2) the fast walking data and (3) combined the data corresponding to slow and fast walking information into a single dataset. This represents a challenging scenario with both dominant identity specific gait information (primary biometric) and non-dominant secondary/soft biometric information (walking style) modelled by LDA/MLP approach.

Table 18: Result of fusion of normal walking and fast walking

<table>
<thead>
<tr>
<th>No</th>
<th>Method</th>
<th>Dataset</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LDA-MLP</td>
<td>Normal Walking</td>
<td>95.5 %</td>
</tr>
<tr>
<td>2</td>
<td>LDA-MLP</td>
<td>Fast Walking</td>
<td>94.5 %</td>
</tr>
<tr>
<td>2</td>
<td>LDA-MLP</td>
<td>Combined</td>
<td>82.5 %</td>
</tr>
</tbody>
</table>

As can be seen in Table 18, while individually fast and slow walking style information modelled by LDA/MLP technique results in good identification accuracy, with 95.5% for normal walking, and 94.5% for fast walking, the modelling of weak soft biometric information (walking style) along with strong biometric information (identity of each subject) weakens the overall identification accuracy (82.5%). However, this depicts more real world scenario, and development of appropriate high performance subspace features and efficient classifier methods can result in better identification performance. It should be noted that the fusion of primary and soft/secondary biometric features is not reported in Table 18 due to lack of space, but some of our preliminary experiments show that fusion of primary and
secondary/soft biometric information (walking style) can result in synergistic fusion. Also, use of motion based static and dynamic features is currently being investigated.

4.9 Multi-Camera Multimodal Fusion
Here I propose an identification approach by considering some of these challenges. Some of the key contributions of the experiment in enhancing the recognition performance are as follows:

- Investigation of feature level fusion of multiple view gait images (3 different views)
- Investigation of feature level fusion of visible and infrared gait images, and
- Investigation of soft/secondary biometric (the walking speed here is the soft biometric).

Next Section describes the details of the datasets, and the proposed identification scheme.

4.9.1 Gait Based Person Identification scheme
For experimental evaluation of our proposed identification scheme, I used CASIA Gait Database collected by Institute of Automation, Chinese Academy of Sciences [52]. It is a large multi-view gait database, which is created in January 2005. There are more than 300 subjects. I used 5 different set of data known as datasets A, B, C, D and E. Dataset A, B and C was captured in 36 degree, 90 degree and 126 degree accordingly with normal video camera. And dataset D and F captured with infrared camera in fast walking and normal walking modes. In total 50 subjects (person) are involved in the experiments. - for all 5 different 5 datasets, these 50 people are commonly involved. In our experiments each person is considered as one class. Hence, each dataset has 50 classes and I used 16 images (corresponding to one gait cycle) for each class. In total four thousand (4000) images were involved in the experiment. The videos were all captured at night.
The five different datasets used (Datasets A, B, C, D and E) provide us with a rich opportunity to investigate different approaches for modelling some of the real world operating scenarios. Our gait based human identification scheme involved investigation of single mode and multi-mode infrared and visible camera images, single mode and multi-mode visible camera images, and inclusion of walking speed as a soft/secondary biometric for enhancing the recognition performance. To the best of our knowledge, this is one of the first works investigating role of soft/secondary biometrics in gait based human identification systems. Figure 48 shows the schematic for the proposed scheme, and extracted silhouette images. I first extracted images from the gait video sequences applied routine image processing operators to subtract the background, and enhance the salient image features, extracted the silhouettes from each frame, and applied some multivariate statistical dimensionality reduction techniques such as principal component analysis (PCA) and linear discriminant analysis (LDA) to extract the feature vectors. As the purpose of this work was investigating the role of multiview gait images, importance of infrared and visible range camera images, and role of soft/secondary biometric in enhancing the robustness and recognition of automatic identification scheme, I decided to use simple PCA and LDA based feature extraction techniques, and devote more time on investigating different learning classifiers as contributions in this area can provide better generalization ability of the scheme, and address mismatch in training and test scenarios, typical of real world operating environments. It turned out that this approach indeed resulted in a satisfactory performance. Figure 49 describes the experimental process.
It was possible to model the 95% identity specific information in around 20 most significant features. The PCA/LDA feature vector sets extracted from each of the CASIA Datasets (Dataset A,B,C,D and E) were used for training different learning classifiers first with standalone single mode features, and with feature-fused multimode features. Some of the classifiers I examined include – the multi layer perception (MLP), and the Sequential Minimum Optimized SVM classifier (SMO).

While MLP Multi Layer Perceptron (MLP) is a feedforward neural network classifier widely used for pattern classification, recognition, prediction and approximation. SMO is a relatively new type of learning classifier. It decomposes the overall QP problem into QP sub-problems, using Osuna’s theorem to ensure convergence. Unlike the other methods, SMO chooses to solve the smallest possible optimization problem at every step. The advantage of SMO lies in the fact that solving for multi instance multipliers can be done analytically. In addition, SMO requires no extra matrix storage at all. There are two components to SMO: an analytic method for solving for the two Lagrange multipliers and a heuristic for choosing which multipliers to optimize. The details of different classifier approaches is presented elsewhere [66]. The detail of the experiments carried out is described in the next Section.
4.9.2 Experimental Result and Discussion
Different sets of experiments were performed to examine the importance of the long range gait information available from multiple views, infrared and visible cameras, and soft/secondary biometric information available from fast walking/slow walking style of the person. Our first set of experiments involved multi-view fusion on three (3) different datasets 36 degrees, 90 degrees and 126 degrees viewpoints. All data captured for this set of experiments involved visible range camera. The results for LDA features with MLP classifier for single mode (single view point) and feature fusion of all three view point images is shown in Table 19.

<table>
<thead>
<tr>
<th>No</th>
<th>Method</th>
<th>Viewpoint</th>
<th>Recognition Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LDA-MLP</td>
<td>36 Degree</td>
<td>98%</td>
</tr>
<tr>
<td>2</td>
<td>LDA-MLP</td>
<td>90 Degree</td>
<td>84.5%</td>
</tr>
<tr>
<td>3</td>
<td>LDA-MLP</td>
<td>126 Degree</td>
<td>88.88%</td>
</tr>
<tr>
<td>4</td>
<td>LDA-MLP</td>
<td>3D View-Fusion</td>
<td>99%</td>
</tr>
</tbody>
</table>

The recognition accuracy achieved with 36 degrees view point is significant higher as compared to 90 degree and 126 degree viewpoints. With 90 degree and 126 degree viewpoints, I obtained 84.5% and 88.88% recognition accuracy. This could be due to better view of frontal face information from 36 degree view point. However, when I fused all the three views (feature fusion with equal weights), I found that the recognition accuracy was 99% significantly higher than each of these view points. This shows that the fusion of multiple views is synergistic and validates our findings that using multiple views of surveillance video footage with long range videos (without detailed face images), it is possible to perform large scale identification with high level of accuracy, using simple subspace features (LDA) and classifier technique (MLP). Such simple approaches can lead to
real time and real world intelligent video surveillance systems - the beginning of a new generation of security systems for public surveillance deployments. The next set of experiments involved use of visible and infrared camera images and cross camera feature level fusion. For this set of experiments I used two (2) different set of data from two (2) different cameras (infrared and visible camera). Exactly same processing steps were applied as in first set of experiments. In total 100 subjects and 1600 images (16 training images per subject) were involved in two different datasets.

Table 20: Recognition accuracy for cross-camera feature fusion

<table>
<thead>
<tr>
<th>No</th>
<th>Method</th>
<th>Dataset/Method</th>
<th>Recognition Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LDA-MLP</td>
<td>Inferred</td>
<td>92%</td>
</tr>
<tr>
<td>2</td>
<td>LDA-MLP</td>
<td>Visible</td>
<td>94.2%</td>
</tr>
<tr>
<td>3</td>
<td>LDA-MLP</td>
<td>Fusion</td>
<td>98.67%</td>
</tr>
<tr>
<td>4</td>
<td>LDA-SMO</td>
<td>Fusion</td>
<td>99%</td>
</tr>
<tr>
<td>5</td>
<td>LDA-SMO</td>
<td>Fusion</td>
<td>68%</td>
</tr>
<tr>
<td>6</td>
<td>PCA-MLP</td>
<td>Fusion</td>
<td>82%</td>
</tr>
</tbody>
</table>

Table 20 shows the results for our 2nd set of experiments. As can be seen in Table 20, the fusion of visible and infrared images turned out to be very promising, with LDA-SMO classifier as the best performer with 99% recognition accuracy. The LDA-MLP comes close second with 98.67% accuracy, and a relatively lower accuracy achieved with PCA based features. However, the fusion (feature fusion with equal weights) of infrared and visible camera images and use of LDA and SMO classifier performs much better. Showing that fusion is synergistic provided a proper discriminatory subspace is identified (LDA subspace) and trained with an appropriate classifier (SMO classifier).
Finally, the third set of experiments involved investigating soft or secondary biometric information, in terms of walking style fast walking and normal walking) for enhancing the recognition accuracy. The walking style data was available for visible camera images for all 50 subjects (persons). That is the data for each person walking in two (2) different ways (fast and normal walking) was used. In the final set of experiments, I examined three different approaches. First, I applied LDA-MLP separately to (1) normal walking data, (2) the fast walking data and (3) combined the data corresponding to slow and fast walking information into a single dataset. This represents a challenging scenario with both dominant identity specific gait information (primary biometric) and non-dominant secondary/soft biometric information (walking style) modelled by LDA/MLP approach.

As can be seen in Table 21, while individually fast and slow walking style information modelled by LDA/MLP technique results in good identification accuracy, with 95.5% for normal walking, and 94.5% for fast walking, the presence of weak soft biometric information (walking style) along with strong biometric information (identity of each subject) weakens the identification accuracy (82.5%). However, this depicts more real world scenario, and development of appropriate high performance subspace features and efficient classifier.
methods can result in better identification techniques. It should be noted that the fusion of primary and soft/secondary biometric features is not reported in Table 21, but some of our preliminary experiments show that fusion of primary and secondary/soft biometric information (walking style) can result in synergistic fusion. Also, use of motion based static and dynamic features is being investigated.

4.10 Summary
In this chapter I proposed a novel multimodal identification approach in different mode of fusion. I did fusion in camera-viewpoint with gait biometric cues form low resolution surveillance videos. The result shows significant improvement in feature fusion of gait cues acquired from different camera viewpoints. I envisage the approach will work better with use of better quality data. However, our proposed multi modal fusion technique is suitable for identifying people in low resolution surveillance videos, which a futuristic and next generation solution for real world identity verification problem. Further, this will allow diffusion of biometric security technologies with better user-acceptability for day to day civilian access control and public surveillance applications. I also, experimented, multi-mode walking feature fusion from low resolution surveillance video for large scale human identification, for which I applied three (3) different dataset captured with infrared camera. The experimental result shows the multi-mode of walking fusion approach worked extremely well, indicating the potential of this approach to real time real world public surveillance applications.

Furthermore, I did fusion based on using different type of datasets based on visible and infrared gait images with side or profile views, and set of feature extraction and classification techniques. Basically I used all data which are in 90 degree view angle. Because of 90 degree view angle, all of our expected traits (ear, side face, and gait) had clear view. Combination of dimensionality reduction approach PCA-LDA with different classifiers I received promising
results. Significant outcome of this experiment is; for surveillance applications infrared camera will work better than normal video camera and that is what I proved by our results.

Finally, I proposed a novel gait based large scale human-identification approach from surveillance videos I performed 13 different experiments to investigate the role of multi view gait images, importance of infrared and visible images, and role of weak soft/secondary biometric for enhancing the recognition accuracy. As can be seen from the experimental results, I can say that use of simple subspace based dimensionality reduction techniques (LDA/PCA), efficient classifier methods (MLP/SMO) and feature level fusion can result in robust gait based identification systems for real world operating scenarios.
Chapter 5: Development of University of Canberra Multimodal Gait (UCMG) Database

5.1 Introduction
Human identity verification from arbitrary views is a very challenging problem, especially when one is walking at a distance. Lately, recognizing identity from gait patterns has become a popular area of research in biometrics and computer vision, and one of the most successful applications of image analysis and understanding. Gait recognition is one of new and important biometric technologies based on behavioral characteristics, and it involves identifying individuals by their walking patterns. Gait can be captured at a distance by using low resolution devices, while other biometrics needs higher resolution. Gait is difficult to disguise, and can be performed at a distance or at low resolution and requires no body-invading equipment to capture gait information. Gait recognition can hence be considered as a powerful recognition technology for next-generation surveillance and access control applications, with applicability to many civilian and high security environments such as airports, banks, military bases, car parks, railway stations etc [67]. Further, gait is an inherently multimodal biometric as proposed in [35], suggesting that there are 24 different components to human gait, and involves not only the lower body but also the upper body motion, including head and the hands. If all gait movements from full body images can be captured, it can be a truly an unique biometric for ascertaining identity. By considering all mentioned potentiality of gait biometric and shortcomings of existing biometric gait-research database; I recently developed a database called UCMG-Database. I strongly believe the database can be used for establishing human identity by using face, side-face, ear and gait. It also can be used for analysis in medical treatment.

Further, The University of Canberra Multimodal Gait Database (UCMG-Database) was developed in University of Canberra in mid-2013. This is the database 1st in Australia of its kind, which is accessible to the researchers in the biometric area around the world (URL: http://staff.estem-uc.edu.au/emdad/). For a single person, it has seven (7) different types of
walking pattern including (i) normal walking, (ii) fast walking, (iii) walking with heavy bag, (iv) walking with overcoat (long jacket), (v) walking with hat, (vi) walking with hoody and (vii) walking with mask. All these sequences have been captured with four (4) different cameras in four (4) different dimensions including; 130 degree, 270 degree, 315 degree view point, and 315 degree from top (roof) with surveillance Camera. More than hundred (100) individual and around three thousand (3000) of different video-sequence contains in the UCBM-Database. The database can be used for person identification based on side-face, frontal face, ear and gait. It also can be used for medical analysis. I’m expecting the database is one of largest and one of unique database (in terms of trait) in the world.

5.2 Literature Search for UCMG-Database
The recognition of people is of great importance, since it allows us to have a greater control about when a person has access to certain information, area or simply to identify if the person is the one who claims to be [2], [37]. And a natural tool to identify a person is the biometric trait. Automated recognition (face) technology first captured the public attention from the media reaction to a trial implementation at the January 2001 super bowl, which captured surveillance image and compared them to a database mug shots [37]. Since then, due to increase in global demand for automated security and surveillance products, there was a proliferation of research works on identity verification based on different biometric modalities [7], [30], and [31]. Several research works have also reported importance of using multiple modalities instead of single biometric trait in order to enhance the accuracy and robustness [30], [31] and [32]. However, most of the systems have two modality, face-gait, face-ear, frontal face, or hand geometry etc. which has been experimented in controlled laboratory environments, and it is a big challenging to achieve similar accuracies and robustness in real world public surveillance applications. Hence, by considering shortcomings and limitations of appropriate database for the research of biometric identification, I
developed an extensive multimodal biometric database. Further, because of uprising demand of robust biometric identity verification system, database is most important part to go. In fact, I have seen only few databases that are available publically for biometric research, even though they have significant limitations. Researcher said; the field of gait analysis is gaining increasing attention for applications such as visual surveillance, human-computer interfaces, and gait recognition and rehabilitation. Numerous algorithms have been developed for analysing and processing gait data; however, a standard database for their systematic evaluation does not exist. Instead, existing gait databases consist of subsets of kinematic, kinetic, and electromyography activity recordings by different investigators, at separate laboratories, and under varying conditions. Thus, the existing databases are neither homogenous nor sufficiently populated to statistically validate the algorithms [68]. It has also limitations on video sequence, data-collection environment too. Table 22 is a comparison on some existing biometric research databases.

Table 22: List of popular database and their modality

<table>
<thead>
<tr>
<th>Name of the Database</th>
<th>Number Video Sequence</th>
<th>Number of Subject</th>
<th>Potential Modality (s)</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gait BaseLine [69]</td>
<td>452</td>
<td>74</td>
<td>Side Face and Gait</td>
<td>The United States</td>
</tr>
<tr>
<td>Human Gait Database [70]</td>
<td>1224</td>
<td>192</td>
<td>Side Face, Ear and Gait</td>
<td>England</td>
</tr>
<tr>
<td>CASIA Gait Database [52]</td>
<td>960</td>
<td>153</td>
<td>Side Face, Gait</td>
<td>China</td>
</tr>
<tr>
<td>Recognizing Human Action [36]</td>
<td>100</td>
<td>25</td>
<td>Side Face and Gait</td>
<td>Sweden</td>
</tr>
</tbody>
</table>


However, as I can see from the above table, gait baseline database [69] contains 452 sequences collected from 74 subjects. It is not suitable for a project which requires ear, side face and gait together, because “Ear” is not clear in the collected video sequence. Any feature or score will not be extractable from ear biometric by using this dataset. Likewise, human gait database [70] seems huge, but some of the dataset will not be useful for a project like multimodal fusion, they used three different approach to collect the data that are; symmetry, Velocity Moments and Moment-Based. In those approaches ear biometric is not clear in most the dataset. In fact, the dataset has basically designed for only-gait analysis.

Moreover, CASIA gait database [52] contains 153 subjects and typically ear was not in concern. It can be applied to only side-face and gait analysis prior to achieve the database from the authority. Further, our objective was to develop a database which will have at least three (3) clear biometric identification traits and that is what I did. With the side face or frontal-face, if I can capture the ear, there will have big potentiality for the system to become more robust.

Furthermore, the database called “recognizing human action” [36] which is publically available. In fact, I have the database and applied the dataset in number of experiments. The subject and sequence are very limited; besides, the ear biometric is not clear in most of the sequences. In total the database contains 25 participants. Finally OSAKA gait database, which is really a big collection of data for ear-side-face and gait research. One noteworthy drawback is; it has been collected completely in-control environment. They also applied specific walking pool for data collection. Even though they have large number of dataset and verity of approach (mode of data-collection), but it will be very week to testify a system for
real-world public surveillance. Therefore, by considering limitations and weakness of existing database I developed UCMG-Database. The specification the database described in the next section.

5.3 MMB Dataset

The University of Canberra Multimodal Biometric (UCMG-Database) Database was developed at University of Canberra, with the approval of “Australian Human-Research Ethics-Committee”, the UCMG-Database has established in April 2013. I designed my project studio in combination of real world CC-camera and digital movie camera. Figure 50 shows the studio set-up and Table 23 shows the data-summary of the UCMG-Database.

![Project Studio](image)

**Figure 50: Project Studio**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of Participant</td>
<td>103</td>
</tr>
<tr>
<td>2</td>
<td>Number of Camera Used</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Number of Video Sequence for each Participant</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>Number of Potential Trait</td>
<td>5 (Frontal face, side face, ear, gait, body movement)</td>
</tr>
<tr>
<td>5</td>
<td>Number of View Points</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>Number of Dataset</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>Number of Sequence in each Dataset</td>
<td>412</td>
</tr>
<tr>
<td>8</td>
<td>Total Number of Video Sequence</td>
<td>2800</td>
</tr>
</tbody>
</table>
Further, the database has seven different models and each model has four (4) different dimensions. In total twenty eight (28) different set of data has nearly three thousand (3000) video sequences. Based on number of trait, number of sequence and number of dimension; I claim this database is one of largest biometric research database in the world. However, the detail of the datasets is as follow;

5.3.1 Normal Walking
In an open studio we have collected the data; one hundred and three (103) different people have been participated. Each participant has four (4) different video pattern based on view point. I have captured from 130 degree, 270 degree 315 degree and 315 degree from top (roof) with CC Camera. All participants walked the way they walk every day. Hence in normal walking platform I have around 412 different video sequence captured by 4 different camera in 4 different view point. Figure 51 shows the sample images of a participant’s normal walking in 130 degree view point.

![Sample images of normal walking](image)

Figure 51: Normal walk in 130 degree view angel.

5.3.2 Fast Walking
In this stand all participants was walking very fast “like running”. Our purpose was to capture the gait while a person walks fast or run after stealing something. In total one hundred and three (103) different people have been participated. Each participant has four (4) different video pattern based on view point. I have captured from 130 degree, 270 degree, 315 degree and 315 degree from top (roof) with CC Camera. All participants walked the way they walk
every day. Hence in this walking platform I have around 412 different video sequence captured by 4 different camera in 4 different view point. Figure 52, shows the sample image of a participant’s fast walking in different view.

5.3.3 Normal Walking with Heavy Bag
To measure the gait variation on weight lifting, I collected data while the participants carried a heavy bag. In this approach researcher can measure overall change in the body movement. Hence altogether one hundred and three (103) different people have been participated. Same as last two (2) sets, each participant has four (4) different video pattern based on view point. And I have captured from 130 degree, 270 degree, 315 degree and 315 degree from top (roof) with CC Camera. All participants walked the way they walk every day. Therefore in this walking platform I have around 412 different video sequence captured by 4 different camera in 4 different view point. Figure 53 shows the sample image of a participant’s walking heavy bag.

![Figure 52: Fast walk](image)

![Figure 53: Walking with heavy bag](image)
5.3.4 Normal Walking with Long Jacket (Overcoat)
The criminal may change over time in external presence, by considering that, I collected data by using long jacket (over coat), in that I can measure the difference between different-walking-approach and probability of detection. For this set of collection, I had same number of people in the same platform. Figure 54 shows the sample image of a participant’s walking pattern by using long jacket (overcoat)

![Figure 54: Walking with long jacket (overcoat)](image)

5.3.5 Normal Walking with Hat
Hat is another approach to examine the identification in compare to walking without hat. In fact, by using hat, person identification based on side-face might be different than without hat. To testify our claim, I collected around 412 video sequences captured by four (4) different cameras simultaneously. I have captured from 130 degree, 270 degree, 315 degree and 315 degree from top (roof) with CC Camera. All participants walked the way they walk every day. In total I had one hundred and three (103) different participants for this approach of our video collection. Figure 55, shows the sample image of a participant’s walking pattern by using hat.


5.3.6 Normal Walking with Hoody
To hide facial lateral, criminal may use hoody as an instrument. By considering the possibility, I collected data with hoody. Again I had one hundred and three (103) different participants. Each participant has four (4) different video pattern based on view point. I have captured from 130 degree, 270 degree, 315 degree and 315 degree from top (roof) with CC Camera. All participants walked the way they walk every day. For this set of data, I have same number video sequence (412) captured by 4 different camera in 4 different viewpoints. Figure 56, shows the sample image of a participant’s walking by using hoody.

5.3.7 Normal Walking with Mask
Robbery by using mask is very common scenario as I can find in crime-report from world media. Our intention is to make a possible way to identify a person even though they use mask. I can now easily measure the difference in the face-ear-gait identification matrix with
mask and without mask. Same number of participant (103) I invited to this part of our video data collection. I also applied same approach in view point with four (4) different (high and low resolution) cameras. Figure 57, shows the sample image of a participant’s walking pattern by using mask.

![Figure 57: Walking with mask](image)

Further, each of mentioned dataset has four dimensions with nearly 412 video sequences. Every single set of data will be investigated five (5) times, including individual experiment of each dimension and the fusion of all (4) dimensions. For seven different set of data I are expecting to have thirty five (35) different experiments by using number of statistical classifiers in combination of dimensionality reduction algorithm.

### 5.4 Summary

In this chapter I have presented multimodal biometric research database (UCMG-Database) which is potentially the largest database in the world of its kind. This dataset (s) can be used for face recognition, facial analysis, side face recognition, ear recognition, body movement analysis, gait recognition, gait analysis etc. It will also be used in the area of medical analysis as well as security analysis for human identification. Further, the database contains nearly three thousand (3000) video sequences captured in seven (7) different patterns (clothing/style etc.) by using four (4) different cameras. It has also a set of data which has captured by a real-time surveillance camera. Hence, I strongly believe, the database going to be useful and effective for the researcher in the area of biometric analysis; for its multimodality, multidimensionality, and real-time operating scenario.
Chapter 6: Multimodal Face-Gait-Ear Fusion with UCMG Database

6.1 Introduction

Over the last few years, several research works have been reported on use of different biometric modalities for establishing identity, and have recognized the importance of gait patterns (walking style), for recognizing human ID at a distance, where the face cannot be clearly seen in video footage. However, these complex surveillance environments are characterized by uncooperative cameras and uncooperative subjects, where most of traditional approaches used for biometric identity recognition with face, iris and fingerprint biometrics normally fail. However, several physiological and biomechanical studies have shown that human gait is a unique and an inherently multimodal biometric, and involves a complex kinematic interaction between several motion articulators, and includes interplay between lower and upper limbs and other biomechanics of joints. It is person specific based on body weight, height, joint mobility in the limbs, and other person specific behavioral nuances. Due to this it is unique and cannot be forged, and if I can model these inherently multimodal traits by extracting compact representations from multiple sources, in terms of robust features and combine them appropriately, it is possible to identify humans from a distance from their gait or from the way they walk, irrespective of uncooperative cameras and uncooperative subjects. Here, even if frontal face is not visible, it is possible to establish the identity of the person using several subtle static and dynamic multimodal cues from frontal and profile face, ear and head shape, walking style and speed, head and hand motion during walking etc.

Automatic identification systems built using these features, can make a great contribution to surveillance and security area, can lead to better understanding of gait abnormalities, and lead to development of better human computer interfaces. However, each of these cues or traits captured from long range low resolution surveillance video cameras on their own are not
powerful enough for ascertaining identity, a combination or fusion of each of them, along with suitable automatic processing approaches can result in robust recognition.

In this chapter, I propose usage of full profile silhouettes of persons from multi-view low resolution cameras for capturing inherently multi-modal cues available from the gait patterns of the walking humans, and use it for establishing their identity. Further, I propose the use of unsupervised feature learning techniques, based on variants of principal component analysis (PCA) and deep learning (DL) approaches, which allow an in-depth analysis of the underlying pixel data. I compare these new features to standard features based on multivariate statistical techniques, such as PCA and linear discriminant analysis (LDA), along with well-known learning classifier approaches based on support vector machines, NN and MLP classifiers [44], [67], [72]. The experimental evaluation of the proposed approach with two different databases, the publicly available CASIA [52] gait database, and the newly developed UCMG database [73], show a significant improvement in recognition performance with the proposed unsupervised learning features, as compared to standard features proposed in the literature, particularly for uncooperative camera conditions, simulated with mismatched train and test data sets. The rest of the chapter is organised as follows. Next Section describes the background and motivation for proposed work, followed by the proposed multiview multimodal feature learning scheme.

6.2 Background
To address the next generation security and surveillance requirements for not just high security environments, but also for day-to-day civilian access control applications with low level security requirements, I need a robust and invariant biometric trait [44]. According to the authors in [7], the expectations of next generation identity verification involve, addressing issues related to application requirements, user concern and integration. Some of the suggestions made to address the requirements of these emerging applications were use of
non-intrusive biometric traits, role of soft biometrics or dominant primary and non-dominant secondary identifiers and importance of novel automatic processing techniques. To conform to these recommendations; often there is a need to combine multiple physiological and behavioural biometric cues, leading to so called multimodal biometric identification system. Each of the traits, physiological or behavioural have distinct advantages, for example; the behavioural biometrics can be collected non-obtrusively or even without the knowledge of the user. While most behavioural biometrics are not unique enough to provide reliable human identification they have been proved to be sufficiently high accurate [74]. Gait, is a similar powerful behavioural biometric, but as a single mode, on its own, it cannot be considered as a strong biometric to identify a person.

However, if I combine complementary gait information from another source, the multi-modal combination is expected to be powerful for human identification. Researchers have found that one of the most promising techniques is the use of multimodality or combination of different biometric traits or same biometric trait from multiple disparate sources. For example, researchers in [40], [75] have found that multi-modal scheme involving PCA on combined image of ear and face biometric results in significant improvement over either individual biometric. In addition, other recent attempts to improve the recognition accuracy include face, fingerprint and hand geometry [18]; face, fingerprint and speech [19]; face and iris [20]; face and ear [21]; and face and speech [22].
However, the power of automatic discovery of suitable feature representations extracted from disparate but complementary sources straight from pixels, that do not rely on elaborate computation intensive features and application-specific expert knowledge, and their fusion did not attract much attention from the research community. As opposed to traditional sophisticated computation intensive feature extraction stages, and use of domain specific expert knowledge to manually specify features, the proposed feature learning and discovery based on the variants of principal component analysis and the deep learning approach, seeks to optimize an objective function that captures the appropriateness of the features, and includes approaches based on energy minimization, manifold learning, and deep learning using auto-encoders [76], [77], [78].

6.3 Methodology
Afresh I applied same method like our previous chapter. On dimensionality reduction and Eigen-value extraction, I castoff principle component analysis (PCA) and linear discriminant analysis (LDA). To identify a person from extracted Eigen value, I castoff most powerful mathematical-function “multilayer perceptron (MLP)”.

6.3.1 Principle Component Analysis
Principle component analysis is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be
hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analysing data. The other main advantage of PCA is that once I have found these patterns in the data, and we can compress the data, e.g. by reducing the number of dimensions, without much loss of information. Basically this technique used in image compression[54]. In the image analysis it works like;

\[ X=(x_1, x_2, x_3 \ldots N_2) \ldots \ldots \ldots \ldots \ldots (18) \]

where the rows of pixels in the image are placed one after the other to form a one dimensional image. Each image is \( N \) pixels high by \( N \) pixels wide. For each image it creates an image vector. And then it counts all the images together in one big image-matrix like;

\[ \text{Matrix} = (v_1, v_2, v_3 \ldots v_N) \ldots \ldots \ldots \ldots \ldots (19) \]

6.3.2 Linear Discriminant Analysis

On the other hand, the LDA also closely related to principal component analysis (PCA) and factor analysis in that they both look for linear combinations of variables which best explain the data. LDA explicitly attempts to model the difference between the classes of data. PCA on the other hand does not take into account any difference in class, and factor analysis builds the feature combinations based on differences rather than similarities. Discriminant analysis is also different from factor analysis in that it is not an interdependence technique: a distinction between independent variables and dependent variables (also called criterion variables) must be made. LDA works when the measurements made on independent variables for each observation are continuous quantities. When dealing with categorical independent variables, the equivalent technique is discriminant correspondence analysis [55].

6.3.3 Multilayer Perceptron (MLP)

Multi Layer perceptron (MLP) is a feed forward neural network with one or more layers between input and output layer. Feed forward means that data flows in one direction from input to output layer (forward). This type of network is trained with the back-propagation
learning algorithm. MLPs are widely used for pattern classification, recognition, prediction and approximation. Multi Layer Perceptron can solve problems which are not linearly separable [57]. In our experiments I had 49 input layer, 800 hidden layer (for each data set) and 50 output layer. This is mostly based on dimensions, instances and the classes of the dataset. The detail of the experiments is described in the next Section.

6.4 Experiment
For experimental evaluation I used UCMG-Gait-Database [73]. The University of Canberra Multimodal Gait Database (UCMG-Database) has developed in University of Canberra in mid-2013. This is the database 1st in Australia of its kind, which is accessible to the researchers in the biometric area around the world. For a single person, it has seven (7) different types of walking pattern including (i) normal walking, (ii) fast walking, (iii) walking with heavy bag, (iv) walking with overcoat (long jacket), (v) walking with hat, (vi) walking with hoody and (vii) walking with mask. All theses of sequences has captured with four (4) different cameras in four (4) different dimensions including; 130 degree, 270 degree, 315 degree view point, and 315 degree from top (roof) with surveillance Camera. More than hundred (100) individual and around three thousand (3000) of different video-sequence contains in the UCBM-Database. In our research laboratory I have examine most the dataset that has described as follows;

For the experiments I have taken dataset from 275 degree view point in number of pattern like; normal walking, fast walking, walking with hat, walking with long jacket, walking with mask, data taken real world CCTV. In total 11 dataset (person) processed for this experiment. In all dataset I have taken same people. Each people has 16 images, therefor I experimented 880 images for this experiment. Table 24 shows the accuracy of identification
### Table 24: Identification with LDA-MLP

<table>
<thead>
<tr>
<th>No</th>
<th>Dataset</th>
<th>Number of Image</th>
<th>Accurate Identification (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Normal walking</td>
<td>880</td>
<td>100%</td>
</tr>
<tr>
<td>2</td>
<td>Fast Walking</td>
<td>880</td>
<td>100%</td>
</tr>
<tr>
<td>3</td>
<td>Walking with Overcoat</td>
<td>880</td>
<td>100%</td>
</tr>
<tr>
<td>4</td>
<td>Walking with Hat</td>
<td>880</td>
<td>100%</td>
</tr>
<tr>
<td>5</td>
<td>Walking with Mask</td>
<td>864</td>
<td>100%</td>
</tr>
</tbody>
</table>

I tried to find out what are the variations if a person changes the walking style, dress and appearance. By using mentioned algorithm and classifier I received equally 100% correction rate in detecting a person accurately. One most ground-breaking finding of this experiment is; to identify a person even they are using mask to cover their complete head. I received 100% correct detection in even if people using mask. Figure 57 and 58 shows the Eigen face and PCA Eigen value.
Subsequently successful 1\textsuperscript{st} experiment, I moved to 2\textsuperscript{nd} experiment with the data extracted from real world CCTV. In total same number of (11) people were involved and equal number of (880) sample image applied to this experiment. 40\% of the data I trained and rest 60\%
used for testing. In this experiment I also received surprising result. Figure 62 showing the result of the data taken from CCTV footage.

As we can see from above figure, I tested 11 people, 16 images each, in total 176 instances. And function MLP identified all 176 correctly. It is also remarkable that true positive and false positive rate 1 and 0 accordingly. However, after getting wonderful from CCTV footage, I moved to my final experiment. In this experiment I tested a person in three clear traits that ear-side face and gait. In total for 11 people 880 images taken for experiment. Out 880 images, 40% used training and rest of them used for testing. Initially I experimented with clear ear and side face. I received 100% correct detection rate, at the end I move to “only-gait”, unpredictably that’s also give me 100% accurate result in identify a person. Figure 63, 64 and 65 shows the Eigen face for side face-ear, Eigen face “gait-only” and threshold curve.
Figure 63: Extracted Eigen Face for Ear-Side Face

Figure 64: Extracted Eigen Face for Gait-Only
6.5 Experiment and comparison with CASIA Gait Database and UCM Gait Database
To benchmark my database I did a comparative experiment with CASIA Gait Database [52] and UCM Gait database [73]. Both the databases are large multi-view databases, and consist of video sequences of walking persons captured from multiple video cameras. Further, both the databases consist of subsets of gait data captured from multiple view angles (0, 30, 60, 90 etc), with different walking styles (slow, normal, fast), and with different props (bag, hat, coat etc.). For all experiments reported in this chapter, I used 10 subjects with a set of extracted silhouettes from Dataset B in CASIA and a dataset with similar matched conditions from the UCMG dataset. Each subject consists of 16 images and in total 160 images for 10 subjects (people). Figure 66 shows some sample sequences from CASIA Gait Database and UCMG Database.
6.5.1 Methodology
In this experiment I applied number of different algorithm and classifier to compare with previously applied classifier and algorithm that has described below.

6.5.1.1 Variant of PCA Feature (v-PCA)
PCA is a basic form of feature learning and it allows automatic discovery of compact and meaningful representation of raw data without relying complex feature extraction techniques or on domain specific (or expert) knowledge. It is a well-established technique used for de-correlation and dimensionality reduction of data. The variance of the original data is concentrated in low dimensional subspace characterized by eigenvectors and eigenvalues. The projection of the original data onto the variance-maximizing sub-space serves as a feature representation, and automatic analysis of the eigenvalue spectrum of the sample covariance uncovers the appropriate target-dimensionality of the feature space. However, the
PCA features perform poorly if the input data are not properly normalized. Using blind range normalization does not solve the problem especially when the components relate to completely different aspects of a phenomenon. In the context of gait recognition from multiple views this becomes problematic and to address this issue I developed an alternate representation based on the empirical cumulative distribution function (ECDF) of the gait silhouette/contour \((x, y)\) from each frame. This representation is independent of the absolute ranges but preserves structural information. For every frame in the gait image sequence, I extract the silhouettes using background subtraction, and extraction of boundary contours, which serves as the gait silhouette. The \(xi\) contour points of silhouette are processed by using an empirical cumulative distribution function ECDF, \(E_i\) of the whitened samples \(\rightarrow x_i\) along each axis \((i=1, 2)\), using standard Kaplan-Meier estimation. These ECDFs, which monotonically increase within the range of zero to one, describe the probability that the \(x, y\) co-ordinates of the silhouette/contour points \(\rightarrow x_i\) are less than or equal to some specific value.

By means of cubic interpolation \(\mathcal{C}^p\), the inverse of ECDF function, \(E_i^{-1}\) at a fixed set of \(N\) points \(p^{-} = \{p_i, \ldots, p_n\}\) is estimated, and this serves as the representation of \(\rightarrow x_i\) of silhouette points. This technique allows normalization of all contour points from images from multiple camera views to a common range, without destroying inherent structural independence (for \(i=1,2\)).

\[
\rightarrow x_i = \mathcal{C}^p (E_i^{-1}(x_i)) \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots
I extracted the reduced dimensionality feature vector for ECDF normalized contour points for the gait silhouettes in each frame of the walking sequences in the datasets using PCA, and we call this as the variant of PCA, or vPCA features.

6.5.1.2 Deep Learning Feature (DLF)

Hinton et al, [76] have proposed a powerful tool for generic semi-supervised discovery of features called auto encoder networks, which aim to learn a lower dimensional representation of input data. This produces a minimal error when used for reconstructing the original data. For auto encoder based feature learning on sequential data I used a novel deep learning approach. In this approach, the lower dimensional features are discovered by means of feed-forward neural networks that consist of one input layer, one output layer and an odd number of hidden layers. Every layer is fully connected to the adjacent layers and a non-linear activation function is used. The objective function during training is the reconstruction of the input data at the output layer. The auto encoder transmits a description of the input data across each layer of the network. This non-linear low-dimensional encoding is hence an automatically learned feature representation. For robust model training, I used the techniques suggested by Hinton et al [76], where the layers of the auto encoder network are learnt greedily in a bottom-up procedure, by treating each pair of subsequent layers in the encoder as a Restricted Boltzmann Machine (RBM). An RBM is a fully connected, bipartite, two-layer graphical model, which is able to generatively model data. It trains a set of stochastic binary hidden units which effectively act as low-level feature detectors. One RBM is trained for each pair of subsequent layers by treating the activation probabilities of the feature detectors of one RBM as input-data for the next. Once the stack of RBMs is trained, the generative model is unrolled to obtain our final fully initialized auto encoder network for feature learning. Different methods exist to model real-valued input units in RBMs. I used Gaussian visible units for the first level RBM that activate binary, stochastic feature detectors.
(Gaussian-binary). The subsequent layers rely on the common binary-binary RBM, and the final layer is a binary linear RBM, which effectively performs a linear projection. During training, the sample data is processed batch-wise, where each batch ideally comprises samples from all classes in the training-set. As the availability of the class information is not mandatory, I trained RBMs in a completely unsupervised manner. To evaluate the two feature learning approaches, \( v^\text{PCA} \) and \( DLF \) features for gait based human identity recognition; I conducted a number of experiments using different subsets of data from the CASIA dataset B and UCMG database. For baseline comparison, I also extract standard PCA and LDA features.

Further, we examined how these features perform with different classifiers and hence tested with Nearest Neighbor (NN), MLP, SVM and SMO classifiers, and has described in detail in some of the previously reported works [15, 16, 17, 19, 24, 27].

6.6 Experiment
Three sets of experiments were performed on different subsets of data from two databases, the CASIA Dataset B (Visible spectrum), Dataset C (Infrared spectrum) and the UCMG database. Table 25, 26 and 27 shows the recognition performance for each set of experiments in terms of recognition accuracy and several statistically significant performance measures such as true positive rate (TPR), false positive rate (FPR), precision, recall and F-measure.

<table>
<thead>
<tr>
<th>Dataset /Feature</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASIA-PCA</td>
<td>48.63</td>
</tr>
<tr>
<td>UCMG-PCA</td>
<td>46.62</td>
</tr>
<tr>
<td>CASIA-vPCA</td>
<td>52.68</td>
</tr>
<tr>
<td>UCMG-vPCA</td>
<td>51.57</td>
</tr>
<tr>
<td>CASIA-DLF</td>
<td>79.5</td>
</tr>
<tr>
<td>UCMG-DLF</td>
<td>78.4</td>
</tr>
<tr>
<td>CASIA-LDA</td>
<td>75.3</td>
</tr>
<tr>
<td>UCMG-LDA</td>
<td>100</td>
</tr>
</tbody>
</table>
By using PCA, LDA, vPCA and DLF feature learning techniques, I extracted the gait feature vectors from the silhouette images of walking humans in each video sequence, and performed identification experiments in single mode and multimodal fusion mode. To examine the performance of features under uncooperative camera conditions, I used different views for training and test conditions. I obtained a fused (averaged) training template by combining features extracted from different views, and used the testing data from a view other than those used for building the fused training template. Without this approach, the error becomes too large if training data is used from one view and test data is used from a different view.

Tables 26 and 27 shows the performance for each set of experiments in terms of recognition accuracy for different sets of features and gait databases. The first set of experiments show the performance for proposed learning features in single mode for CASIA dataset B and UCMG datasets, under matched conditions (gender, walking styles and presence of props). For all experiments I used a reduced dataset, with 10 classes (10 persons) with 20 features (PCA, vPCA, LDA, DLF), as preliminary experimentation showed that about 95% of variations can be modelled by around 10 features, and any more increase in dimensionality does not result in significant improvement in performance. As we can see, in Table 25 the LDA and DLF performing very well.

The second set of experiments involved testing whether other established classifiers with different kernels result in better performance, and is shown in Table 26. As can be seen in Table 26, for both datasets, simple nearest neighbour classifier outperforms other sophisticated SVM classifier for all kernel types, which could be due to better learning ability of DLF features.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Database</th>
<th>Accuracy (%)</th>
<th>TP</th>
<th>FP</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>CASIA</td>
<td>92.5</td>
<td>0.93</td>
<td>0</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>NN</td>
<td>UCMG</td>
<td>92.25</td>
<td>0.92</td>
<td>0</td>
<td>0.93</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>SVM-L</td>
<td>CASIA</td>
<td>81.13</td>
<td>0.81</td>
<td>0</td>
<td>0.82</td>
<td>0.81</td>
<td>0.81</td>
</tr>
</tbody>
</table>
The final set of experiments involved examining the decision level fusion of LDA and DLF features for different types of classifiers. As can be seen in Table 27, the decision fusion of UCMG dataset with LDA-MLP performing extremely well (100% accuracy) than the single mode features, particularly for NN and SVM classifier with linear kernels.

### 6.7 Summary

In this chapter I examined my developed UCMG database. For dimensionality reduction and feature extraction, I applied principle component analysis (PCA) and linear discriminant analysis (LDA). To identify a person from extracted feature I applied a classifier called “Multilayer Perceptron (MLP)”. In combination with all three, I received 100% accuracy in correctly identification of person. Furthermore, to compare the accuracy and robustness of two datasets, I have developed a novel protocol for involving comparison of developed “UCMG-Database” and well known CASIA-Database. As can be seen from the experiments described in this chapter, UCMG is giving more accurate results in comparison with CASIA-database. In fact, both datasets give us poor results for the first experiment for this chapter. I identified; this is because of feature extraction algorithm. In first experiment (UCMG-only) I applied, PCA and LDA for dimensionality reduction and feature extraction, but for the new protocol developed based on comparison of UCMG and CASIA, which I call as
“comparison-experiment” I applied vPCA and DLF instead. Hence I can summarize the chapter that, LDA and MLP is the best combination to identify person from low resolution surveillance video.
Chapter 7: Conclusions and Further Work
With my thesis I have established a novel human-identification scheme from long range face-gait profiles in surveillance video. I investigated the role of multi view face-gait images acquired from multiple cameras, the importance of surveillance and visible range images in ascertaining identity, the impact of multimodal fusion, efficient subspace features and classifier methods, and along with side face-ear; the role of soft/secondary biometric (walking style) in enhancing the accuracy and robustness of the identification systems.

Experimental evaluation of several subspace based side face-ear, gait feature extraction approaches and learning classifier methods on different datasets from publicly available databases (CASIA-China, Human Action Database- Sweden and UCMG Database- University of Canberra) shown significant improvement in recognition accuracies with multimodal fusion of multi-view face-ear, gait images from visible and infrared cameras acquired from video surveillance scenarios.

Further work involves, working on approaches similar to DL approaches I have examined in this thesis. The use of DL or Deep Learning approaches remove the involvement of human experts in feature engineering, or crafting of or design of features manually. Deep Learning approach is an automatic unsupervised learning approach, and there are not many approaches proposed based on unsupervised learning for large scale human identity recognition in real world complex surveillance scenarios with seamless continuous stream capture of video signals from multiple cameras, and a need to perform real time identity recognition in airport, immigration and border control check points. Here, in these scenarios, it is not optimal to process the video footage offline, with carefully handcrafted mathematically intense feature sets. What is needed is automatic real time processing of raw video data with minimal human involvement. The use of deep learning features proposed here, offers a significant promise, and further development of the approaches similar to deep learning, and those that are based
on unsupervised and semi-supervised learning can lead to significant breakthroughs in automated security and surveillance technologies.
Bibliography


