HUMAN IDENTIFICATION USING GAIT

A thesis submitted in partial fulfilment of the requirement for

M.Tech Dual Degree

in

Electronics and Communication Engineering

(Specialization: Communication and Signal Processing)

By

Abhijit Nayak

Roll No: 710EC4048

Department of Electronics and Communication Engineering

National Institute of Technology Rourkela

Rourkela, Odisha, 769008, India

May 2015

HUMAN IDENTIFICATION USING GAIT

A thesis submitted in partial fulfilment of the requirement for

M.Tech Dual Degree

in

Electronics and Communication Engineering

(Specialization: Communication and Signal Processing)

By

Abhijit Nayak

Roll No: 710EC4048

Under the Guidance of

Dr. Samit Ari

Department of Electronics and Communication Engineering

National Institute of Technology Rourkela

Rourkela, Odisha, 769008, India

May 2015

DEPT. OF ELECRTONICS AND COMMUNICATION ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY
ROURKELA, ODISHA -769008

CERTIFICATE

This is to certify that the work presented in the thesis entitled **Human Identification using Gait** by **Abhijit Nayak** is a record of the original research work carried out by him at National Institute of Technology, Rourkela under my supervision and guidance during 2014-2015 in partial fulfilment for the award of Dual Degree in Electronics and Communication Engineering (Communication and Signal Processing), National Institute of Technology, Rourkela.

Place: NIT Rourkela

Date:

Dr. Samit Ari

Assistant Professor

DEPT. OF ELECRTONICS AND COMMUNICATION ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY
ROURKELA, ODISHA -769008

DECLARATION

I hereby declare that the work presented in the thesis entitled **"Human Identification using Gait"** is a bonafide record of the systematic research work done by me under the general supervision of **Prof. Samit Ari**, Dept. of Electronics and Communication Engineering, National Institute of Technology, Rourkela, India and that no part thereof has been presented for the reward of any other degree. I also declare that due credit has been given to information from other sources wherever used in this work through citations and details have been given in references.

Abhijit Nayak
710EC4048

ACKNOWLEDGEMENT

This research work is partly made possible because of the continuous motivation by so many people from every part of my life. I convey my deepest regards and sincere thanks to my supervisor Prof. Samit Ari for his esteemed direction and support throughout the period of this research work. I would also like to thank all the faculty members of the Department of Electronics and Communication Engineering, NIT Rourkela for their valuable help in the completion of my thesis work. I extend my gratitude and sincere thanks to the fellow students and research scholars at the Pattern Recognition Lab, Dept. of ECE for their constant motivation and cooperation throughout the tenure of this work. Finally, I would like to express my sincere thanks to my family and friends for their unending encouragement and sharing of experiences, and without whose support this work could never have been completed successfully.

Abhijit Nayak
abhijitnayak25@gmail.com

INDEX

ABSTRACT

Keeping in view the growing importance of biometric signatures in automated security and surveillance systems, human gait recognition provides a low-cost non-obtrusive method for reliable person identification and is a promising area for research. This work employs a gait recognition process with binary silhouette-based input images and Hidden Markov Model (HMM)-based classification. The performance of the recognition method depends significantly on the quality of the extracted binary silhouettes. In this work, a computationally low-cost fuzzy correlogram based method is employed for background subtraction. Even highly robust background subtraction and shadow elimination algorithms produce erroneous outputs at times with missing body portions, which consequently affect the recognition performance. Frame Difference Energy Image (FDEI) reconstruction is performed to alleviate the detrimental effect of improperly extracted silhouettes and to make the recognition method robust to partial incompleteness. Subsequently, features are extracted via two methods and fed to the HMM-based classifier which uses Viterbi decoding and Baum-Welch algorithm to compute similarity scores and carry out identification. The direct method uses extracted wavelet features directly for classification while the indirect method maps the higher-dimensional features into a lower-dimensional space by means of a Frame-to-Exemplar-Distance (FED) vector. The FED uses the distance measure between pre-determined exemplars and the feature vectors of the current frame as an identification criterion. This work achieves an overall sensitivity of 86.44 % and 71.39 % using the direct and indirect approaches respectively. Also, variation in recognition performance is observed with change in the viewing angle and N and optimal performance is obtained when the path of subject parallel to camera axis (viewing angle of 0 degree) and at N = 5. The maximum recognition accuracy levels of 86.44 % and 80.93 % with and without FDEI reconstruction respectively also demonstrate the significance of FDEI reconstruction step.

LIST OF ABBREVIATIONS

ATC : Air Traffic Control
CASIA : Chinese Academy of Sciences
CCTV : Closed Circuit Television
CGEI : Clusteral Gait Energy Image
CMC : Cumulative Match Characteristic
DEI : De-noised Energy Image
DWT : Discrete Wavelet Transform
EM : Expectation Minimization
FDEI : Frame Difference Energy Image
FED : Frame to Exemplar Distance
FHMM : Factorial Hidden Markov Model
FOV : Field of View
FPS : Frames per Second
GEI : Gate Energy Image
GHI : Gait History Image
GMI : Gait Moment Image
GMM : Gaussian Mixture Model
GPLVM : Gaussian Process Latent Variable Model
HMM : Hidden Markov Model
IDTW : Improved Dynamic Time Warping
IPD : Inner Product Distance
LDA : Linear Discriminant Analysis
MEI : Motion Energy Image
PCA : Principal Component Analysis
PHMM : Parallel Hidden Markov Model
SBHMM : Segmentally Boosted Hidden Markov Model

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Human Gait as a Biometric

Gait, in simple terms, refers to the walking style or motion style of an individual or entity. The individual could be an animal, a human or even a robot. The gait of a person can present various cues about the individual, including information about age, sex, physical disabilities, identity etc. The human brain has evolved to recognize persons by seeing their gait. Thus, just like face recognition, where we can identify a person by seeing his face, we are also able to identify a person just by looking at the style of movement. It is important to note that gait

It is apparent that if humans can recognize other humans by their gait, a computer vision system can also recognize humans by recording their gait signatures if the gait data of the person is already present in the system. This work is primarily concerned with the use of human gait for the purpose of automated person recognition.

In recent years, keeping in view the escalating security threats and commission of anti-social/malicious acts around the world, increasing attention is being given to the effective identification of individuals. Trends in the development of security and surveillance systems, especially automated ones, reflects the growing importance of biometrics. Biometrics are more reliable measures of identification compared to human-defined identification measures like ID numbers, cards etc. as they are inherently less vulnerable to duplication and faking. Features for biometric signatures are selected such that there is minimal (if not zero) probability of identical signatures being generated by any two random subjects. Essentially, this guarantees a lesser probability of unintentional mis-identification.

There are many biometric features such as fingerprint, iris detection, face detection, for which algorithms have already been developed. For some of these, the detection accuracy is satisfactory enough for practical use in ‘reasonably-controlled’ real-time environments. Compared to these techniques, gait recognition, until now, has reached lower levels of correct identification levels. Also, it has been tested on databases that are primarily generated in highly controlled environments that raises questions over their applicability in real-time scenarios where a lot of factors can affect the identification process, starting from lighting conditions to change in the direction of motion of subject and accessories carried by subject (bags, winter clothing etc.) to occlusions (self-occlusions as well as occlusions by other individuals in public places).

For practical scenarios, the present status of gait is that it can only be used a secondary biometric, along with some primary biometric signature that has a better reliability of detection and correct

identification. In these scenarios, gait is used as an initial screening mechanism to monitor and dispense the majority of the cases, while the cases that need further monitoring are referred to the primary biometric.

In-spite of all these shortcomings, human gait has shown to be a promising biometric and research work in this field is going on because of the following positive attributes

1.1.1 Advantages of gait as a biometric

1. **Non-obtrusive or Non-invasive**: Gait is a non-obtrusive technology, which means that active coordination of the subject is NOT required for the collection of information. For fingerprint detection, the subject is required to press his finger on a predesignated surface so as to let the system retrieve his signature. For iris detection or face detection, the subject is required to position his body in a certain predesignated angle/position so as to let the camera capture the required details. The beauty of non-invasive techniques such as gait recognition is that the subject is not required to perform such tasks. For example, is the gait recognition system is meant for biometric-based attendance purposes, it can be simply fitted at the entrance door. As the subject walks in, the camera can capture the details and start the processing, without invading the user.
2. **Maintaining Secrecy**: Compared to other biometrics, gait signatures can be captured more secretively. First of all, they do not intrude upon the subject, so the subject has no way of knowing whether he is being monitored or not (provided the camera is concealed suitably). Second of all, gait data can be taken from a fairly large distance, so it is less conspicuous in nature. Thirdly, thermal cameras can be used for night-time surveillance even without ambient lighting, so the system can be concealed in dark environments that need monitoring.
3. **Lesser Image Resolution Required**: Resource-wise, gait is relatively less demanding compared to other imaging-based systems like face recognition, because the image resolution needed for gait identification is lower. A normal camera feed at a modest frame rate of, say 25 fps, may be installed for reliable identification. Most importantly, even the entire gamut of information of the human body is not required, most algorithms employ just the silhouette/outline of the human body for identification. Thus, in situations where

the image data is of insufficient quality, gait will fare as a more robust system compared to other image-based biometrics.

4. **Circumstantial advantages/disadvantages**: Many biometrics have certain inherent weaknesses based on specific circumstances that make them vulnerable to breaches. For example, if face detection systems are employed in ATM booths, malicious intruders can simply wear masks while entering (since they know it's a sensitive location and CCTVs would be in place). This simple measure can negate the face detection security system completely. Similarly, if camera systems are not around, fingerprint detection systems at isolated places can be breached by exploiting the body of unconscious/dead authorized personnel to gain access. But it is very difficult to duplicate the gait of another person, and hence gaining illegal access by breaching a gait identification system would be virtually impossible. However, to evade identification, while you can't change your biometric signatures like retina or fingerprint, walking style can be altered slightly to dupe the system. This is possible only if the subject has prior information about the installation of a gait recognition system.

1.2 Common parameters in gait analysis

Gait Cycle:

A gait cycle represents the fundamental temporal unit of processing in gait recognition, and corresponds to a periodic cycle that transits from Rest to Right-Foot-Forward (RFF) to Rest to Left-Foot-Forward (LFF) to Rest position [1]. This basically encompasses the entire range of possible positions that a human body passes in the overall course of walking. Fig. 1.1 shows the five characteristic positions of a typical human gait cycle.

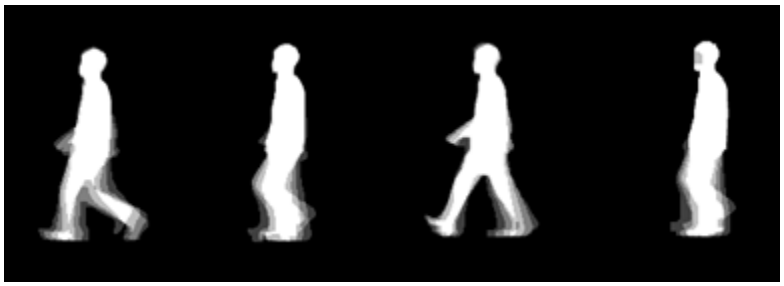

Fig. 1.1 Characteristic positions of a typical human gait cycle

As the subject walks across the Field-of-View (FoV) of the camera, multiple gait cycles are usually captured depending upon the FoV of the camera and the gait dynamics of the individual. The gait

cycle, being the basic unit in gait-based image processing, contains information about dynamic motion and relative motion among all the body parts as the individual moves. In other words, the dynamics and periodicity of the gait cycle characterizes the motion of the individual, along with the static features like height, width etc. The gait cycles are repetitive in nature and as the number of acquired gait cycles increases, there is a consequent increase in information redundancy.

Gait Period/ Gait Cycle Period:

In simple terms, gait period is the time required for a person to complete one gait cycle. But, since the camera records image information in the form of frames, and the frames are periodic in nature (e.g. 30 frames-per-second or 30 fps), it is more useful to obtain the gait period of a person in terms of frames. This can be easily obtained by finding the number of frames elapsed between the starting and ending frames of an extracted gait cycle.

Gait period in itself can be used as an identification feature, but if the fps is low, the gait period of most individuals falls within a narrow range. As a result, its discriminatory power decreases, and it can be used only in tandem with other more discriminatory features.

Gait period also gives us cues about the speed of the person. For example, if the average gait period for a particular system is 28 frames, and an unknown individual takes 45 frames for one gait cycle (gait period = 45), then it can be deduced that the speed of the person is markedly slow compared to the norm. This can be particularly useful for systems that use gait to determine age, as old persons are more probable to walk at lower speeds.

Stride Length:

It is the maximum stretch between the limbs of a person, and is a potential feature for identification. Its value can be measured by placing a bounding box around the individual in the image. Stride length value is obtained from multiple gait cycles and the result is averaged in order to make it more robust to noise and miscalculations.

1.3 Literature Review – Approach to the Identification problem

The existing approaches in image processing and computer vision dealing with the problem of gait identification fall into two broad categories – model-based approaches and model-free approaches.

1. **Model-based approaches:** These methods assume a-priori models to represent gait and match the 2-D gait image sequence to the model parameters. They obtain a series of static/dynamic features by modelling various portions of the body and the manner of their

motion. Once the matching is accomplished, feature correspondence is achieved and is used for recognition. Chen et al. [2] propose a new representation called FDEI and use it as a previous step of HMM-based recognition while Xue et al. [3] use infrared gait data with Support Vector Machines (SVM) for identification. [4] and [5] propose background subtraction and exemplars-based HMM respectively for human tracking and activity recognition. Lee et al [6] have used a model-based approach where ellipse-fitting is used to represent the 2-D images in terms of several ellipses and the geometrical parameters of ellipses are used for characterization and recognition of gait sequences. Model-based features utilize both static and dynamic parameters from bodily features, and generally exhibit angle (view) and shift (scale) invariance. Cunado et al. [7] have matched thigh movement to an articulated motion model, and thus use the hip-rotation angle as an identification characteristic. The primary problem of model-based approaches is that they are dependent on the quality of the silhouette images.

2. **Model-free approaches:** In this approach, there is no pre-assumed model. Instead, successive frames are used to predict/estimate features related to shape, velocity, position etc. These features are calculated for all the persons in the database and are subsequently used for identification. Huang et al [8] have used optical flow as a parameter to characterize the motion sequence in a gait cycle and Principal Component Analysis (PCA) to derive eigen-gaits which are used as discriminating features. Little et al [9] have used features based on frequency and phase extracted from the optical flow parameters of the image. Template matching is carried out to perform recognition.

In this project, a model-free approach has been employed to carry out gait identification. The general framework of automatic gait recognition (using model-free approaches) consists of person detection, extraction of binary silhouettes, feature extraction, and classification stages. After the detection process (which determines whether a subject is present in the current frame), there is a need to discard unnecessary information that is not required for the identification process. Background subtraction is widely used for this purpose in order to separate the individual from the image background, which is achieved by using the difference between a background model (which is updated after every frame) and the current frame [5]. Background subtraction and extraction of binary silhouettes can be treated as pre-processing steps in the entire process.

After this step, feature extraction is a crucial step for effective identification. It investigates and determines the features that can be exploited for recognition, which are subsequently extracted from the silhouette image sequences. There exist a variety of model-free features based on the use of only binary silhouettes and there is no need for the construction of any model to represent the gait dynamics of the subject [9] [10] [11]. The features extracted from segmented video sequences possess high dimensionality and are generally not effective for direct use in the recognition process. Also, a high degree of redundancy is encountered in these feature vectors. Consequently, dimensionality reduction methods are employed to suitably represent these feature vectors in lower-dimensional space. Many such methods are proposed in the literature, among which Principal Component Analysis (PCA) [10] and Linear Discriminant Analysis (LDA) [12] have been most popular.

Classification stage marks the final stage of the entire identification process. This consists of finding the subject whose gait characteristics are most likely to match with the gait characteristics of the subject in the test sequence. Thus, this is a probabilistic measure, and involves the ranking of all the individuals in the training database according to the degree of matching with the test subject. The highest ranked individual determines the identity of the unknown test subject. Feature classification in gait generally employs three approaches or methods. The first method is direct classification, which is generally used after a single template representation or extraction of key points/frames from the gait sequence. The second method employs the degree of similarity between temporal gait sequences to quantify and measure a distance feature, which is then used to estimate the probability as to how closely the test sequence is represented by any random training sequence. The case with the lowest value of distance measure identifies the test subject. In contrast to the above two methods, the third method employs state-space based modelling such as Hidden Markov Models (HMM) [13]– [17]. This approach is primarily focused on the pattern of transition between various pre-defined states related to succession of stances in a temporal gait sequence. This approach employs the similarity criterion between probe and training data as well as the shape appearance [2]. For this reason, the third approach has been employed in this work.

For the extraction of binary silhouettes, background subtraction is a commonly employed method. Segmentation methods involving background subtraction [16][17] and optic flow models [18][19] to find the coherent motion are common.

However, even the most robust background subtraction methods involve exceptions and anomalies and produce erroneous results at times which are detrimental to the recognition performance. There can be many factors that can result in imperfect segmentation of the human body from the background. These include similar intensity levels of the foreground (person) and the background elements, abrupt changes in illumination, occlusion or moving objects in the foreground or background, variation in the distance or viewing angle between the camera and the subject, etc. As a result, there is occurrence of noise elements and spurious pixels, artifacts or bright spots, shadow elements, holes inside the moving silhouette, and missing body portions – all leading to imperfect silhouettes. This is true even if the acquired image sequence is of relatively good quality. These incomplete or partially correct silhouettes may affect the recognition performance significantly. Thus, it is imperative that in order to make the recognition robust to these abnormalities, these low-quality binary silhouettes need further processing. Small defects like noise elements or small holes can be removed by common morphological techniques such as erosion and dilation. However, if the scale of imperfection or incompleteness is higher, for example, missing entire body parts, specific algorithms aiming at reproducing the silhouettes need to be applied.

These algorithms can be broadly classed into three approaches –

Silhouette Reconstruction: Liu et al. [20] and Liu and Sarkar [21] tried to reconstruct the silhouettes using Hidden Markov Models (HMMs). HMM is used to create a mapping from the gait frame sequence to a particular exemplar or stance. Subsequently, the silhouette reconstruction is performed by means of an appearance-based model. The advantages include robustness of the silhouettes to variation in viewing angles and orientation, however the characteristic information contained in a single image is generally lost, thereby affecting recognition performance.

Contour Alignment: These methods work by aligning the contours of the silhouettes of adjacent frames in a sequence as it is assumed that the imperfect extraction will affect only a small number of frames and not the entire gait sequence, which is true in most cases. Yu et al [22] proposed an Improved Dynamic Time Warping (IDTW) to deal with occurrence of noise elements in subject silhouettes or contours by aligning each point on one contour to several points on another by means of conventional Dynamic Time Warping. All pairs except the one with the shortest distance are discarded. The problem with this approach is that it is more vulnerable to undesirable results when

the silhouette imperfection exists for the entire gait cycle. However, this is rare and since most methods use data from neighboring frames, they are susceptible to failure in this scenario.

Enhancing robustness in static representation: In this method, the gait cycle is compressed into a set of one or more static images. The recognition performance would then depend on the quality of these static images. Han and Bhanu [15] proposed a static representation, the Gait Energy Image (GEI) that encompasses both static and temporal information, and is computed by simply taking the mean of features of all the centre-aligned silhouettes of the given gait sequence. GEI is found to be relatively less susceptible to noise effects in individual frames when the noise at different moments is uncorrelated, but the amount of temporal information contained is very low and most of the information is static information. A few representations based on the GEI were developed later, and include the Gait History Image (GHI) [23] and Gait Moment Image (GMI) [24]. The GHI preserves the dynamic or temporal information to some extent, but the primary shortcoming is that there exists only one GHI for each gait cycle. Since the number of gait cycles in the database are limited, this creates the problem of limited number of image sets to train the classifier. GMI, on the other hand, represents the probability at the key moments of all gait cycles. A number of stances/positions or 'key moments' are pre-defined, and the frames corresponding to these key moments in all the gait cycles are averaged to obtain the respective moment GEIs. This overcomes the limitation of small number of training images as encountered in GHI, but the chief issue in the case of GMI is the selection of the key moments. Since all the gait cycles for a particular subject do not always have the same gait period, it becomes difficult to select the key moments by assigning a temporal index.

This work uses the FDEI representation as proposed in [2] which falls under the third category of static representations. This representation represents both static and temporal information satisfactorily and there is one FDEI image per frame, so it is not limited by the number of training sets available. It alleviates the problem of imperfect silhouettes to a large extent. The details of the FDEI representation and steps of the algorithm have been presented in Sec. 2.1.3.

After the background subtraction and FDEI reconstruction, feature extraction is the most crucial step. Sarkar et al. proposed a baseline algorithm [12] directly uses silhouette images as features. Bobick and Davis [25] propose two static representations of gait data - the Motion Energy Image (MEI) and Motion History Image (MHI) in the form of 2-D signal templates incorporating the information of the gait sequence. Liu et al. [26] determine the GEI-wise contribution in the

classification process. Wavelet features obtained from the GEIs are applied to infrared gait identification by Xue et al. [27]. As mentioned before, Gait Energy Image (GEI), proposed in [15], uses a single 2-D template for representing the entire information of a gait cycle. FDEI [2] is the sum of the GEI and the positive portion of the difference between adjacent temporal frames.

Kale et al. [28] propose contour width of the binary silhouette as a feature, which is defined as the horizontal distance or the number of pixels along the x-axis between the left and right ends or extremes of the binary silhouette. However, for low-resolution gait imaging, taking the silhouette itself is more suitable. The width feature and entire silhouette are both used by Kale et al. later in [1]. Weiming et al. [29] propose the transformation of silhouette contour to a 1-D signal by taking the pixel-to-pixel distance along the silhouette contour and silhouette centroid. The shortcoming of 1-D signals is that they are found to be highly susceptible the quality of silhouettes. Dadashi et al. [30] propose the use of wavelet features extracted from these one-dimensional signals. Boulgouris et al. [31] propose the segmentation of the binary silhouette into several angular sectors in the spatial domain and use the distance measure between the foreground pixels and the centroids of these sectors as a discriminating feature. Weiming et al. [32] analyze the shape of silhouettes using Procrustes shape analysis and a mean shape measure is used as the feature. Boulgouris et al. [33] process the silhouettes using Radon Transform to obtain recognition using 2-D template matching.

Regarding state-space model representation for classification, HMMs representing the various phases of gait motion as hidden states have been widely used. The advantage of using HMM-based approaches over others is that they incorporate both shape similarity features and the temporal relation between shapes, i.e. the manner of succession of frames. HMM has been shown to be robust due to its statistical nature. HMM-based recognition has already been used for speech and gesture recognition [34][35]. Aravind et al. [36] use a generic HMM based method for gait recognition. Kale et al. [28] use a low dimensional 1-D vector, called the FED vector to carry out identification and then used wavelet feature in a direct approach in [1]. Debrunner et al. [37] use Hu moment feature vector sequence and HMMs while Yin et al. [38] extract the most discriminative feature for HMM-based classification by proposing a Segmentally Boosted Hidden Markov Model (SBHMM) to map gait data to a new feature space in a non-linear fashion. Heng et al. [39] construct the Factorial HMM and Parallel HMM having multilayer structures. Cheng et

al. [40] apply Gaussian Process Latent Variable Model or GPLVM to map the gait sequence to lower-dimension and extract motion data (temporal information) using HMM.

Liu et al. [41][42] employ a population HMM to model a pre-defined set of subjects. The generic stances and silhouette sequences are taken as the hidden states and observations respectively and the training is performed on a set of silhouettes specified manually.

There are certain assumptions that have been employed in this work

1. The camera's location remains static, hence the Field of View (FoV) is constant – This is true for most practical scenarios. The camera does not need to move/rotate by tracking the person. It just captures the image from a fixed position and relays the information to the PC/Server.
2. The person walks only along a fixed path – thus the angle between the camera axis and the walking path remains constant. We have taken an angle of 90 degrees for the same. This is true for practical scenarios where there is a narrow pre-defined path in front of the camera and perpendicular to it. But for public places where people walk at different angles to the fixed camera, this assumption does not hold true.
3. Occlusion-free data: We have assumed that the gait sequence of the subject can be obtained without any occlusion (self-induced or occlusion by other objects/individuals). Thus, at any particular instant, there is only one individual in the FoV. This simplifies the analysis to a great extent. Most databases created for gait recognition have been created with this assumption. This assumption is violated in situations where there are a number of persons moving together in a public place. But for situations like gait-based biometric attendance where there is controlled environment and only one person crosses the entrance at a time, this assumption holds true.

Any activity, like walking (gait) is generally comprised of two components:

a) a ***structural component*** that includes factors such as stride length, height of individual, etc.

b) a ***dynamic component*** that includes dynamic information. Dynamic information encompasses any information that accrues because of motion itself, such as the manner of swinging of arms, the manner of change of distance between the lower limbs, etc.

In this project, a systematic approach integrating and incorporating both structural and dynamic information has been used for the aforementioned objective. The process involves three broad steps – pre-processing, feature extraction, and HMM-based gait recognition.

The details of the entire procedure and the methodology are sequentially explained in Chapters 2 and 3. Fig. 1.2 sequentially presents a broad outline the entire methodology.

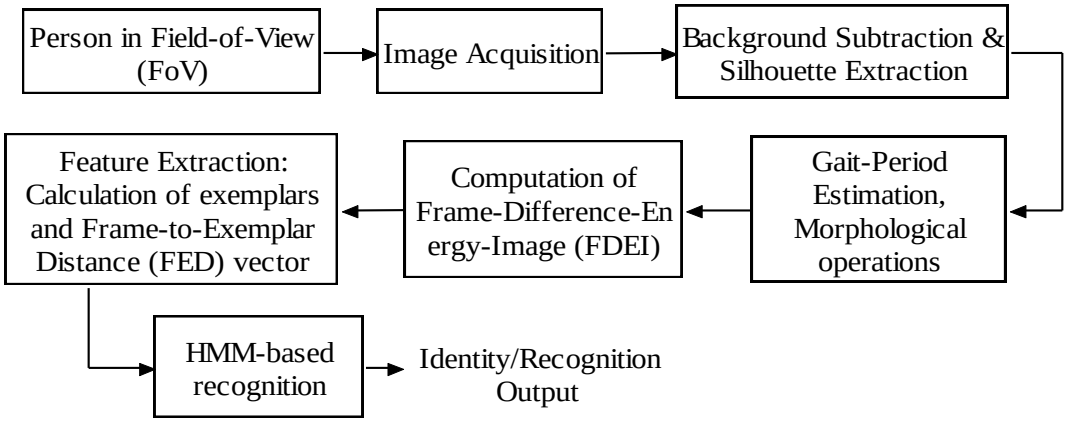

Fig. 1.2 – Broad outline of the gait identification process

1.4 CASIA Gait Database

The CASIA Gait Database is created and provided to promote research in gait recognition by the Institute of Automation at Chinese Academy of Sciences (CASIA). The database consists of four datasets (A, B, C and D) catering to different types of acquisition of gait sequences under varying conditions. This work uses Dataset B of the CASIA Database, which is a larger (compared to A, C, D datasets) multi-view dataset containing the gait data of 124 subjects captured from eleven different viewing angles. This dataset has been used and cited by many research papers [2][13][27][31].

1.5 Motivation

Security & Surveillance techniques are acquiring increasingly greater significance in today's world. These are crucial for routine monitoring, avoiding unauthorized access, detection of security breach, authentication of identity for authorized personnel, etc. Automated systems are becoming pivotal in ensuring 24X7 security and also for other institutions/purposes where human identification is required. They are cost-effective in the long run, and don't get worn out by monotonous work for infinitely long periods. They also rule out the margin for human errors and

negligence which is always a possibility in conventional systems. Thus, there is an immediate need to employ automation at all base levels with minimal manual control at higher levels.

Person identification is an indispensable part of modern surveillance systems as it provides selective access to premises/ facilities. Also, in case of detection of a breach, it helps in zeroing in on the possible suspect. Apart from surveillance, human identification is also used for purposes like registering daily attendance for employees/students/personnel at workplace, academic institutions, and sensitive locations such as Air Traffic Control (ATC) Towers, where there is a pre-defined number of persons who enjoy authorized access.

Biometric signatures like gait are considered reliable for identification systems due to minimal probability of duplication. Human gait analysis research has shown promising results for more extensive use in automated identification systems. Aforementioned factors provide the basis for this work. In this work, a systematic approach to silhouette-based gait identification is performed.

1.6 Thesis Outline

The rest of this thesis is organized as follows. Chapter 2 presents the detailed methodology involved in the pre-processing and feature extraction steps, and the motivation therein. Chapter 3 provides a basic introduction to Hidden Markov Models (HMM) and details the manner of their use in the entire training and recognition process. The observations and experimental results of the recognition method are presented and discussed in Chapter 4, along with variations in performance noted with change in parameters. Finally, the work is concluded in Chapter 5 and future work is reported.

CHAPTER 2

PRE-PROCESSING AND FEATURE EXTRACTION

2.1 Pre-processing

Pre-processing is performed on the acquired images so as to optimize them for feature extraction. This process involves getting rid of redundant information, and maximizing the relevant information. In this work, pre-processing sequentially involves the steps of background subtraction, morphological operations, gait period estimation, and FDEI reconstruction which are described in detail below.

2.1.1 Background Subtraction

Background information present in the Field-of-View (FOV) of the camera is included in the acquired frames, but is not useful for the identification process. For identification, only the static and dynamic information contained in the silhouette of the human subject figure is required. Thus, background subtraction and extraction of silhouettes constitutes the first crucial pre-processing step. As in all pre-processing algorithms, this algorithm should not be computationally extensive so as to increase the time taken for entire process but at the same time it should be efficient enough to produce acceptable results. This work employs a fuzzy-correlogram based method [5] for background subtraction.

Before applying the method, it is important to underline that for gait recognition, only the silhouette of the subject is needed. Thus the output image should be a binary image with the outline of the human subject. All other features of the subject, such as colour of clothing, is irrelevant, since the identification has to rely on gait or motion data only. Thus, this step takes greyscale images as input and produces binary silhouette images as output.

The d-distance correlogram $cor_d(m,n)$ computes the probability with which two given intensity values m and n occur at a distance of d pixels in the given image, and is given by the probability

$$cor_d(m,n) = P(f(x_1) = m,\ f(x_2) = n \mid \|x_1 - x_2\| = d) \tag{2.1.1.1}$$

Thus, a correlogram captures the spatial relation between a pair of pixels in addition to the intensity information. Since taking all the 256 intensity levels individually increases the complexity, grouping the intensity range into *l* bins ($l << 256$) reduces the correlogram size to *l X l*. But a regular correlogram involves crisp assignment of bins and is vulnerable to quantization noise. For example, due to slight illumination changes or artifacts or other errors leading to localized intensity

changes with steep gradients, a particular pair of pixels may contribute to the neighboring bins instead of the actual bins where they should belong. To alleviate this problem, a fuzzy membership function is introduced into the correlogram in [5] to create a Fuzzy Correlogram such that each pixel pair contributes to every bin with a finite and definite probability, while having major belongingness or maximum probability in the adjacent bins.

In the fuzzy correlogram, the membership matrix M is obtained by employing fuzzy c-means algorithm. Also, lesser number of bins are used (c) as compared to the regular correlogram (l^2) which leads to a further reduction in computational complexity. Since it is region-based, the fuzzy correlogram based background subtraction method performs well in case of dynamic backgrounds too.

The sequential steps employed in the background subtraction algorithm are briefly mentioned below.

Step 1: Using Fuzzy c-means algorithm, a c-dimensional fuzzy correlogram vector F is obtained by using a membership matrix M with dimensions $c \; X \; l^2$ and a correlogram vector C with dimensions $l^2 \; X \; l$ as

$$F = M.C \tag{2.1.1.2}$$

M is computed once and remains the same throughout the entire process.

Step 2: The intensity range of the input image is quantized into l levels, where l << 256. In this work, l = 8 has been used.

Step 3: For the first image of sequence, i.e. the frame at time t = 1, for each pixel, compute C taking a window of 8 X 8 pixels around it.

Step 4: Using Eq. 2.1.1.1, for each pixel, compute the Fuzzy Correlogram Vector F. F is taken as the initial background model for the corresponding pixel [5].

Step 5: For all other images, (i.e. t=2, 3, …, N), the Current Fuzzy Correlogram vector is obtained in a similar way and a modified form of K-L divergence distance measure between the Current Model (F) and Background Model is computed as follows

$$D_{KL} = \sum_{i=1}^{c} F_i^c \times \log(\frac{F_i^c}{M_i}) + F_i^b \times \log(\frac{F_i^b}{M_i}), \qquad where \; M_i = \frac{F_i^c + F_i^b}{2} \tag{2.1.1.3}$$

and subscripts b and c indicate background model and current model respectively.

Step 6: For a particular pixel, if this distance measure is less than an empirically determined threshold T, it can be concluded that the current correlogram is reasonably close to the background model and hence the pixel is classified as belonging to background. Thus the pixel is classified as

$$background\ \ pixel, \quad if\ \ D_{KL} < T, \quad \text{and} \quad foreground\ \ pixel, \quad if\ \ D_{KL} \geq T$$

Step 7: The final step updates the background model at each pixel. This is done by replacing the existing background model with the current fuzzy correlogram after adaptive filtering as

$$F_i^b(t) = (1-\alpha)F_i^b(t-1) + \alpha F_i^c(t) \tag{2.1.1.4}$$

where α is the learning rate parameter. If α = 0, it means the background modelling function at time t is same as the one at (t – 1). On the other hand, if α = 1, it means that the background modelling function at t is defined by the fuzzy correlogram vector at time t, and not at all by the background at (t – 1). These are the two extremities, and in this work, an empirically determined α value of 0.01 [5] has been used.

Fig. 2.1 illustrates a sample result of the background subtraction algorithm with a dynamic background and static object in the foreground.

Fig. 2.1 – Sample result of fuzzy correlogram-based background subtraction method

2.1.2 Gait Period Estimation

Gait period estimation is required for two purposes – to use gait period as a feature itself, and to separate gait cycles for further processing. Gait cycles represent the fundamental unit of human gait, and every processing attribute, for example, exemplars, clusters, HMM parameters are defined in accordance with gait cycles. But when a camera captures a moving person, it just captures a stream of digital frames. To group this stream of frames into distinct gait cycles, it is crucial to have a reliable estimate of the gait period.

A host of methods can be used to perform this task – but it is important that this process remains as less time-consuming as possible. A slight deviation does not affect the recognition process drastically, so it is ideal to choose a method that's not highly computationally intensive but at the same time is capable of producing reliable results. In this work, a simple method described in [1] has been used. During any walking cycle, the following two situations are routinely encountered.

- Situation-1: When the legs of a walking subject are stretched to the maximum, i.e. when the distance between both the legs is maximum, the area under non-zero pixels is maximum.
- Situation 2: Conversely, the area under non-zero pixels is minimum when the legs cross each other.

Since walking is a quasi-periodic process, this means that the number of non-zero pixels periodically increases and decreases repeatedly as a person walks. This information is used by the described method to estimate the gait period.

After the completion of background subtraction, the bottom half of each binary silhouette in the input sequence is selected and the number of non-zero (white) pixels are counted. These values are stored in a 1-D vector and plotted. The plot appears as a series of valleys and peaks, with the peaks representing Situation-1 and valleys representing Situation-2. Any one gait cycle involves two peaks and three valleys [Rest (valley) to Right-Foot-Ahead (peak) to Rest (valley) to Left-Foot-Ahead (peak) to Rest position (valley)]. An estimate for the gait period can be obtained by measuring the distance between the first and third valleys. Generally speaking, this can be obtained by measuring the distance (number of frames elapsed) between any two valleys (or peaks) that have one valley (or peak) between them.

Fig. 2.2 illustrates the plot of the vector discussed above. The peaks and valleys represent the maximum and minimum separation between limbs respectively.

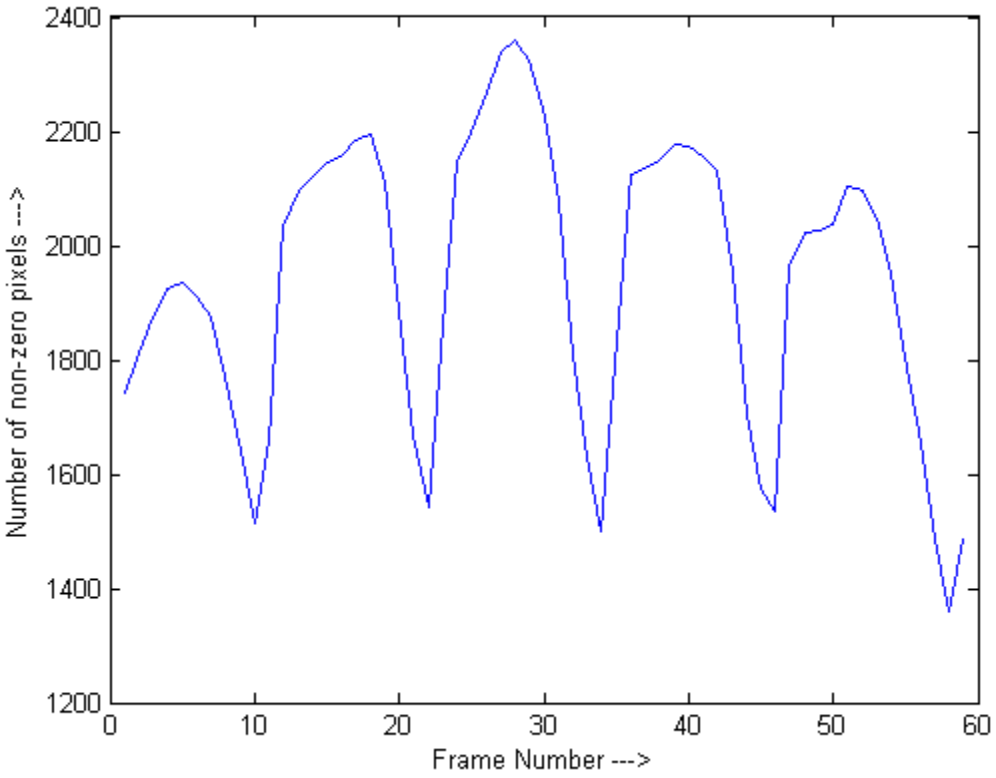

Fig. 2.2 – Sample Plot of number of non-zero pixels in the bottom half of a silhouette over the progression of a gait cycle

2.1.3 Frame Difference Energy Image (FDEI) reconstruction

This final pre-processing step aims to make the recognition process robust to imperfections in silhouette extraction. The quality of extracted human silhouettes is directly related to and crucial for robust gait identification [2]. Often due to exceptions and operational errors in pre-processing algorithms such as background subtraction, incomplete silhouettes are obtained. These imperfect silhouettes present a major problem, since incompleteness of silhouettes appears to be more harmful and intractable compared to other errors such as the presence of noisy artifacts, and can drastically affect recognition performance. To alleviate these effects, this work employs Frame Difference Energy Image [2] to reconstruct the silhouettes and make the recognition process robust to imperfect silhouettes.

The primary motivation for this step is to retain the shape features of the silhouette while mitigating the detrimental effects of imperfect silhouette extraction. The following steps outline the construction of FDEI representation of a gait cycle.

Step 1: Segment the gait cycle into N temporally adjacent clusters and calculate the clusteral GEI or CGEI, which is the mean of all the frames of the particular cluster. The CGEI is a broad representative of the stance of a particular cluster, and is calculated as

$$CGEI(x,y) = \frac{1}{N_C} \sum_{t \in C} f(x,y,t)$$

(2.1.3.1)

This concept is the same as that of GEI which involves the same process for the entire gait cycle, and was first employed in [15]. Here, C refers to the particular cluster, B(x, y, t) refers to the binary silhouette or frame at time t and N_C refers to the number of frames in that cluster.

Fig. 2.3 illustrates the clusteral GEIs of a sample gait cycle. The gait cycle is divided into N temporally adjacent clusters, the number of frames being nearly equal in each. Thereafter, the centroids are aligned and the GEIs of these clusters are generated. The basic stances of human bipedal motion can be seen through these clusteral GEIs as the subject transits through the gait cycle.

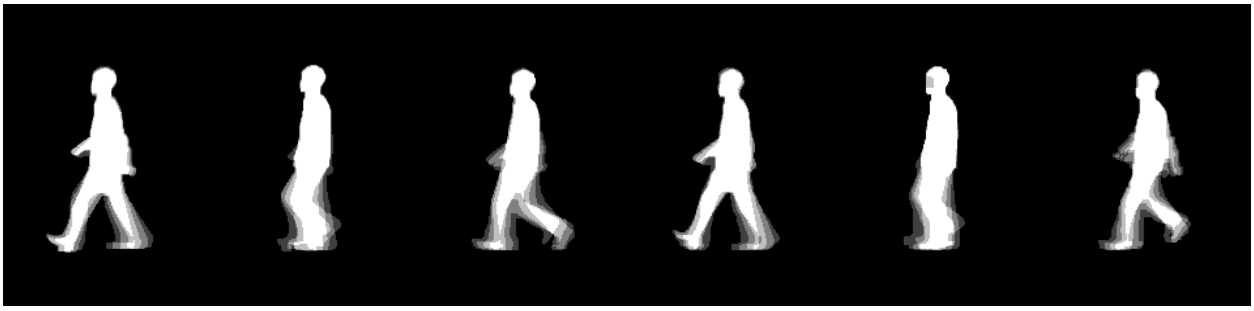

Fig. 2.3 – Clusteral Gait Energy Images (CGEIs) of sample gait cycle

Step 2: This step involves the de-noising of the clusteral GEI by means of a simple thresholding operation based on an empirically selected threshold. This empirically selected threshold is variable, and varies with the change in subjects or gait cycles, and is dependent on the quality of extracted silhouettes. As discussed in [2], the quality of silhouettes is not predictable, and so on the basis of average quality, an experimental threshold is selected as 0.8*max(CGEI), where max(CGEI) denotes the maximum intensity level present in the Clusteral Gait Energy Image.

The de-noising is performed by means of a simple operation as follows

$$D_C(x,y) = \begin{cases} CGEI(x,y), & if \;\; CGEI(x,y) \geq T_C \\ 0, & otherwise \end{cases}$$

(2.1.3.2)

Here, $D_C(x, y)$ is the de-noised CGEI, and T_C is an empirical threshold. Basically, this operation reduces the pixels that are less than T_C to zero, and retains the remaining pixels.

Step 3: This step involves the calculation of 'positive portion' of frame difference. Now, frame difference at time t is defined as the pixel-wise difference between the frames at time instants t and (t - 1). The frame at time t, B(x, y, t) is subtracted from the frame at time (t-1), i.e. B(x, y, t - 1). The positive portion of this frame difference at time t, i.e. $FD(x, y, t)$ is obtained by simply assigning zero to the negative values [2]. Thus, the positive portion of frame difference is defined as follows

$$FD(x, y, t) = \begin{cases} 0, & if \;\; B(x, y, t-1) \le B(x, y, t) \\ B(x, y, t-1) - B(x, y, t), & otherwise \end{cases}$$

(2.1.3.3)

Step 4: This is the final step which involves the construction of the Frame Difference Energy Image at time t, denoted as FDEI(x, y, t), and is defined as the sum of the positive portion of frame difference $FD(x, y, t)$ as obtained in Step 3 above and the de-noised CGEI or $D_C(x, y)$ as obtained in Step 2 above. Thus

$$FDEI(x, y, t) = FD(x, y, t) + D_C(x, y).$$

(2.1.3.4)

There can be two possible cases of incompleteness of silhouettes.

- Case 1: The current frame $B(x, y, t)$ is incomplete while the preceding frame $B(x, y, t-1)$ is complete. In this case, the incomplete portions of the silhouette are contained in $FD(x, y, t)$ and hence, accounted for in the FDEI(x, y, t).
- Case 2: Both $B(x, y, t)$ and $B(x, y, t-1)$ are incomplete. This is the worst-case scenario, and the positive frame difference can't help here. $D_C(x, y)$ may compensate the missing portion to some extent.

In conclusion, the FDEI suppresses the effect of missing parts and makes the imperfect silhouette more complete by preserving its original characteristics. The FDEI is computed for every frame at all the time intervals [2] and it contains the dynamic information (movement part), thereby partially compensating for the missing portions of the extracted silhouettes.

Fig. 2.4 illustrates the FDEI reconstruction process. The FDEI has substantially alleviated the incompleteness of the silhouette, thereby reducing its effect on the recognition process.

Fig. 2.4a – GEI, DEI and positive portion of frame difference

Fig. 2.4b – Original imperfect silhouette, temporally adjacent silhouette, FDEI.

2.2 Concept of Exemplars

During every gait cycle, a set of certain distinct stances or positions can be identified, such as (sequentially) 1- Rest, 2- Hand raised, 3- Hands and Feet Separated, 4- Maximum Displacement between limbs, 5-Return to rest [1]. These stances are generic in nature and each person transits through these over the gait cycle. The information contained in these stances are different for different people, both statically and temporally, and thus can be used as a discriminatory feature. Features corresponding to these position-points are taken as exemplars. It is important to note that exemplars are not images themselves but feature vectors which correspond to these stances.

The motivation for using an exemplars-based method is that recognition can depend on some *distance measure* between the observed silhouette and the exemplars [1].

In this work, the HMM parameters (A, B, π) and the exemplars together represent the identity of a given individual. During training, the exemplars for a particular person are updated after every gait cycle.

In practice, the gait cycle is divided into N temporally adjacent segments, and the initial estimate for the ith exemplar is obtained by and taking the mean of the feature vectors of all the frames included in the ith cluster. The basis for this is the assumption that a group of frames around each generic stance contains the features that represent the stance reasonably well. The update procedure for the exemplars is mentioned in Sec. 3.3.

There are N number of exemplars, which is the same as the number of hidden states in the HMM. The selection of the number N is optimal when the average distortion noise for that value of N is minimum. The problem of picking the optimal value of N is the same as deciding an optimal dimensionality for any stochastic model in order to fit in a given set of observable variables. There are many available methods available for choosing the degree of polynomial regression, analysis of rate distortion curves being one of them [1][4]. In this case, the average distortion is computation depends on the number of exemplars and N is chosen such that the rate of fall in distortion value is appreciably low when the number of exemplars is more than N. It is observed that average distortion falls rapidly up to N = 5, but after that the rate of fall slows down. Thus, N is chosen as the optimal number of exemplars for this case.

2.3 Feature Extraction

Regarding the manner of incorporation of features in the whole process, two approaches are employed – direct approach and indirect approach. In the direct approach, the feature vector is fed directly to the classifier, whereas in the indirect approach, the multi-dimensional image feature vector is mapped on to a lower dimensional space (one-dimensional) and this new 1-D vector is used for the recognition process [1]. The detailed methodology is described below.

2.3.1 Indirect Approach

In this approach, N number of stances are picked from the gait sequence to act as exemplars, and the whole sequence and recognition process is based on this set of exemplars $\varepsilon = \{e_1, e_2, ..., e_N\}$. This N also defines the number of hidden states on which the HMM-based recognition process is based. The selection of the number of exemplars N is done as mentioned in Sec. 2.2.

The primary characteristic of the indirect approach is that the higher dimensional feature vectors extracted from the binary silhouette images of the gait cycle are not directly used in the classification process. Instead, they are mapped on to or represented in a lower-dimensional space

which retains most of the information relevant for classification while reducing redundancy and computational complexity.

The Frame-to-Exemplar-Distance (FED) vector [1] is a measure of reducing the higher dimensional features to lower dimension. Let $f(t)$ represent the feature vector extracted from the binary silhouette image at time t, and $\varepsilon = \{e_1, e_2, ..., e_N\}$ represents the set of exemplars for the current gait cycle. Now, since exemplars are of the same length as individual feature vectors, inner dot product (IDP) can be taken as a distance measure. The distance values of $f(t)$ from the exemplars of the gait cycle constitute the FED vector, such that the distance between the feature vector of the current frame $f(t)$ and the ith exemplar gives the ith entry of the FED vector. This can be represented as

$$[F_j(t)]_i = d(f^j(t), e_i^j) \quad , \; i \in \{1, 2, ..., N\} \tag{2.3.1.1}$$

Where $[F_j(t)]_i$ represents the ith entry of the FED vector computed for the frame at time t in the gait cycle of the jth person, $d()$ represents the distance measure, and e_i^j represents the ith exemplar of the gait cycle of the jth person.

Now, $i \in \{1, 2, ..., N\}$ as in Eq. 2.3.1.1, and distance measure is a scalar value, which suggests that the size of the FED vector will be [1 X N]. This vector, denoted as $F(t)$ acts as a lower dimensional representation for the gait image at time t. Such $F(t)$s are computed for every frame of the gait observation sequence.

Note that there is not one person but a large number of persons, say P persons, whose gait data has to be integrated into the training and recognition system. For training purpose, the exemplars of the ith person is used to compute the FED vector from the frames of the ith person. But in the recognition process, a given set of unknown observations will be available, and FED vectors will have to be computed by taking the distance measure from the N exemplars of all the P persons. To accommodate this, a better way of representation of the FED vector will be $F_j^p(t)$, which denotes that the FED values have been computed by taking the distance between frame features f(t) of jth person and the exemplars of the pth person. Similarly, $[F_j^p(t)]_i$ is used to denote the ith entry of this vector. When $p = j$, i.e. $F_j^j(t)$, it denotes an observation vector of person j. On the other hand,

when $p \neq j$, i.e. $F_j^p(t)$, it denotes the encoding of the gait data of the ith person using the exemplars of the jth person.

As a gait cycle progresses, the distance of the current frame from the exemplars changes [1]. For example, at the beginning of a gait cycle, i.e. for the first frame, it is more likely to be closer to the first exemplar than the remaining four exemplars (assuming N = 5). But as the gait cycle progresses, the distance between the first exemplar and the current frame will increase, and that between the second exemplar and the current frame will decrease, till the distance becomes minimum. After that, the distance between the current frame and the second exemplar will also start increasing again, and the frame will gradually move closer to third exemplar, and so on. Thus, there will be a succession of valleys temporally corresponding to the FED vector.

Most importantly, the FED vector is virtually independent of the choice of features [1] or dimensionality of feature vectors extracted from the observed sequences.

The FED vector can be seen as the observed manifestation of the transition across exemplars or stances (a hidden process) [1]. The whole process can be seen as a Markov process, with exemplars representing the hidden states, and an HMM can be used to model the statistical characteristics of the process according to the observed FED vectors. Thus the FED vectors represent the observation symbols of the HMM. The recognition process is described in detail in Sec. 3.3.

2.3.2 Direct Approach

In this case, the entire feature vector in high-dimensional space is used for the recognition process. Learning and updating the observation symbol probability matrix B is a crucial issue in training.

Wavelet approximation features are shown to represent the most relevant information for person detection [2]. Therefore, a 2-D Discrete Wavelet Transform using Haar wavelet is applied on the FDEIs and the first level approximation coefficients are extracted as feature vectors and are used for further processing. These wavelet vectors are normalized and resized as 1-D vectors to compute their distance from exemplars. Although the overall characteristics of these feature vectors appear similar on a global scale, there are minute variations as illustrated in Fig. 2.5, which are the key to classification and are accounted by HMM.

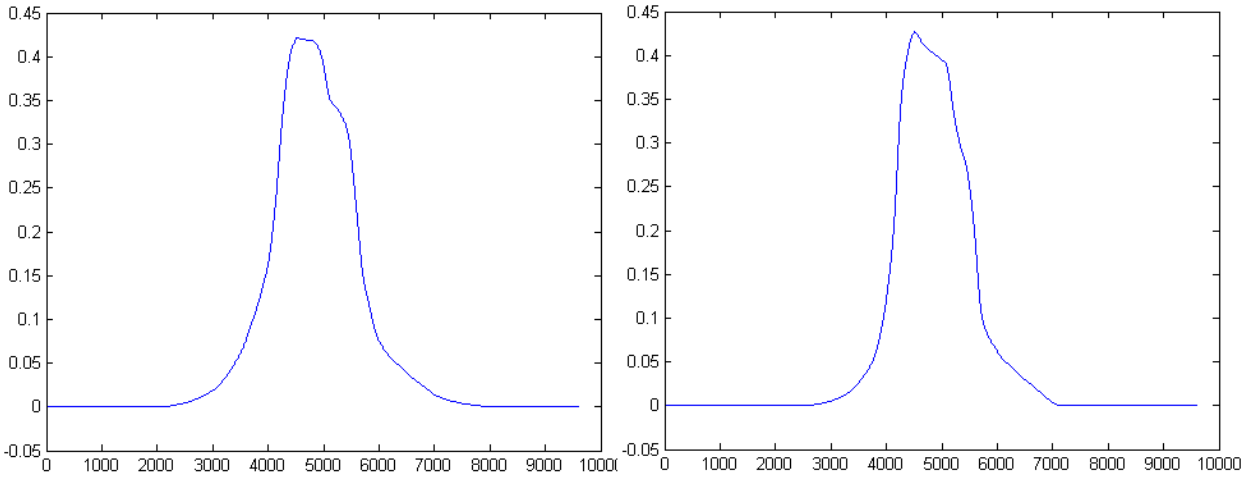

Fig. 2.5 – Visual comparison of feature vectors obtained from gait sequences of two subjects

Because the feature vector is high-dimensional in nature, B can be represented in a modified form as presented in [1]. This alternative representation is based on the distance of feature vector from the exemplars (Frame-to-Exemplar-Distance or FED) as follows:

$$b_i(f(t)) = P(f(t) \mid e_i) = \beta e^{-\alpha D(f(t), e_i)} \tag{2.3.2.1}$$

where $f(t)$ is the frame at time t, e_i is the ith exemplar, and $P(f(t) \mid e_i)$ denotes the probability of observation $f(t)$ being generated by the ith hidden state or exemplar. $D(f(t), e_i)$ represents the distance of current feature vector f (t) from the ith exemplar e_i.

But in this case, the FED is not used as a vector representation of the image itself to be used for classification. Instead, the FED values (or distance values) are used just for defining the observation symbol probability matrix B. This is the significant difference in approach compared to the Indirect Approach described in Sec. 2.3.1. The training and recognition process using the direct approach is described in detail in Sec. 3.3.

2.4 Conclusion

This chapter describes the pre-processing and feature extraction steps of the work. Pre-processing includes three sub-steps – a Fuzzy Correlogram-based background subtraction algorithm followed by gait period estimation and FDEI computation. The background subtraction efficiently distinguishes between the foreground and static/dynamic background while FDEI computation is

shown to significantly alleviate the effect of silhouette imperfection by adding positive frame difference to incompletely extracted silhouettes. Feature extraction step consists of two approaches – direct and indirect, that use high-dimensional wavelet feature vectors and low-dimensional FED vectors respectively. A figurative comparison between feature vectors shows distinguishable patterns which are the key to HMM-based classification. The results incorporating performance accuracy with direct and indirect approaches are presented in Chapter 3. . The choice between the two types of features is primarily guided by a trade-off between computational complexity and recognition accuracy. Use of direct features provides better accuracy levels but at a higher computational cost.

CHAPTER 3

RECOGNITION USING HIDDEN MARKOV MODELS

3.1 Markov Property and Markov Process

Markov Property

In probability theory, Markov Property is said to be satisfied when a stochastic process is memory-less by nature, i.e. when the probability distribution of the future state of the process (conditional on the past and present states) is dependent solely upon the present state of the process and not on the preceding state. Markov assumption is a term that describes a model where it is assumed that the Markov Property holds true, such as a Hidden Markov Model (HMM).

Markov Process

A Markov Process is used to refer to any stochastic process or model that satisfies the Markov Property. Broadly speaking, a process is said to satisfy Markov property if it is possible to predict the future state of the process using the information of the present state only. This means that even if the past history of the entire process is employed for prediction, the prediction will be the same as the one made by looking solely at the present. In other words, the future of the system does not depend on the past states (independent of them) provided the present state of the system is precisely known and is used to predict the future. This is essentially a First Order Markov Process.

Markov processes can be used to model random processes that change states according to some underlying transition rule depending only on the present state. Generally, a Markov Process has a finite or countable state space, or a set of values which a process can take. For example, {rainy, sunny, cloudy} etc. can be states used to model the weather of a particular place.

Order of Markov Process

In a given Markov process, the past states represent a context for determining the probabilities of future states. The number of past events employed by the process to make this prediction is called its *order*. In a first order Markov process, the probabilities for the next future state depends only on the immediately preceding state, or the present state, as described above. Similarly, in a second order Markov process, the future state depends on the last two states, i.e. the present state and the state just preceding it. In a similar fashion, a given Markov process can use any number of past states for prediction, including the degenerate case of no past choices, i.e. a zero order Markov Process. This special case (zero order Markov Process) essentially means that the conditional probability of the future is independent of both the past and the present, which means it is equivalent to a weighted random selection.

Since the choice of number of past states influence the predictability of the future states depending on the nature of the process, a Markov process can model different degrees of variation based on different patterns of data. The higher the order, the closer the process comes to matching the specific pattern it models. The pattern on which a Markov process is based can be determined by a statistical analysis of data. In this work, we model temporal gait sequence as a First-order Markov process. This is because.

Markov Model

Any model used to characterize a Markov process is called a Markov Model. Hidden Markov Models (HMM) are a subset of Markov Models where the state of the process is partially visible or 'hidden'.

3.2 Introduction to Hidden Markov Model

Hidden Markov Model (HMM) is a statistical tool that is employed for modelling a wide range of data involving temporal sequences.

'Hidden' States: As opposed to a general Markov Model (any model used to represent a Markov process), in case of a Hidden Markov Model, the state of the process is only partially visible. Thus, a part of the process is 'hidden' in nature. In other words, a set of observations that are related to the underlying state of the process are visible, but these 'visible observable states' are generally not sufficient to determine the process state precisely. Several algorithms have been established for Hidden Markov Models, which are subsequently discussed in Sec 2.4. Thus, in a general HMM, the states of the process are 'hidden' but they are related to some observable states by means of some underlying principle particular to that process.

Thus, it is clear that there exist two kinds of states – hidden states and observable states. By looking at the observable states, and by understanding the nature of their relation to the hidden states, it is possible to predict or estimate the sequence of hidden states and thereby, to characterize the Markov process. An HMM is developed solely for this purpose, i.e. to 'mimic' the Markov process whose states are partially hidden and to predict its behavior. It is used for modeling systems involving temporal sequences as observations that are characterized by an underlying process.

An HMM framework can be used to model stochastic processes where

I. State of the system is a Markov process and is Non-observable.
II. Observable sequences of system have an underlying probabilistic dependence.

3.2.1 A Basic Example of HMM

To illustrate the above concept, let us consider a simple example of a stock market index with three possible states –

i. Bullish (when the stock market is poised positively suggesting rising demands)
ii. Bearish (when the stock market is poised negatively suggesting falling demands) and
iii. Even (when the market is poised evenly and there is no significant sign of growth or slowdown)

Here, the stock market is assumed to follow a Markov Process. Thus, it is assumed that the future state of the market (whether it will be bullish/bearish/even) can be determined solely by looking at the current state.

The three states described above are non-observable or 'hidden'. However, though we cannot definitively know these states, but these states can be related to and characterized by the changing trends in share prices, which are fully observable. The share index prices show three trends – rising, falling and unchanged. These trends, are, in fact, the 'observable states/sequences' while the state of the market is 'hidden or non-observable'. An HMM can make use of the nature of relation between observable states (change in share prices) and hidden states (nature of stock market) so as to form an approximate model of the process, i.e. the stock market.

Fig. 3.1 shows the Markov process described above along with the transition probabilities.

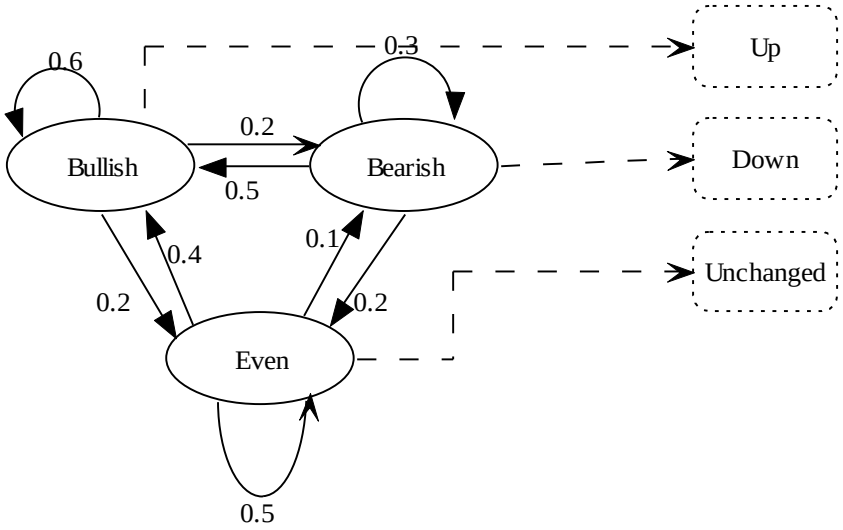

Fig. 3.1 – Representation of a sample First order Markov Model without hidden states

This is a first order Markov Process, since the next state depends only on the current state and the fixed probabilities. To create an HMM to model the stock market index, the states and observations are as following (Note: In this case, number of hidden states (N) = number of observable states (M) = 3. However, in general, it is NOT necessary to have M = N: in most real scenarios $M \neq N$)

Hidden States (nature of market)	Observable States (trends in share prices)
1. Bullish 2. Bearish 3. Even	1. Up 2. Down 3. Unchanged

This process is illustrated as an HMM in Fig. 3.2. Although we can't observe the state of the market, we can observe the current trends in share prices. The objective is to use this observable information in order to characterize the various aspects of the Markov process (stock market). This HMM now allows all observable symbols (trends in share prices) to be emitted from each state with a finite probability. What this means is that a bullish market would have both good days (share prices up) and bad days (share prices down); however, the probability of a being a good day will be more, as shown in the figure.

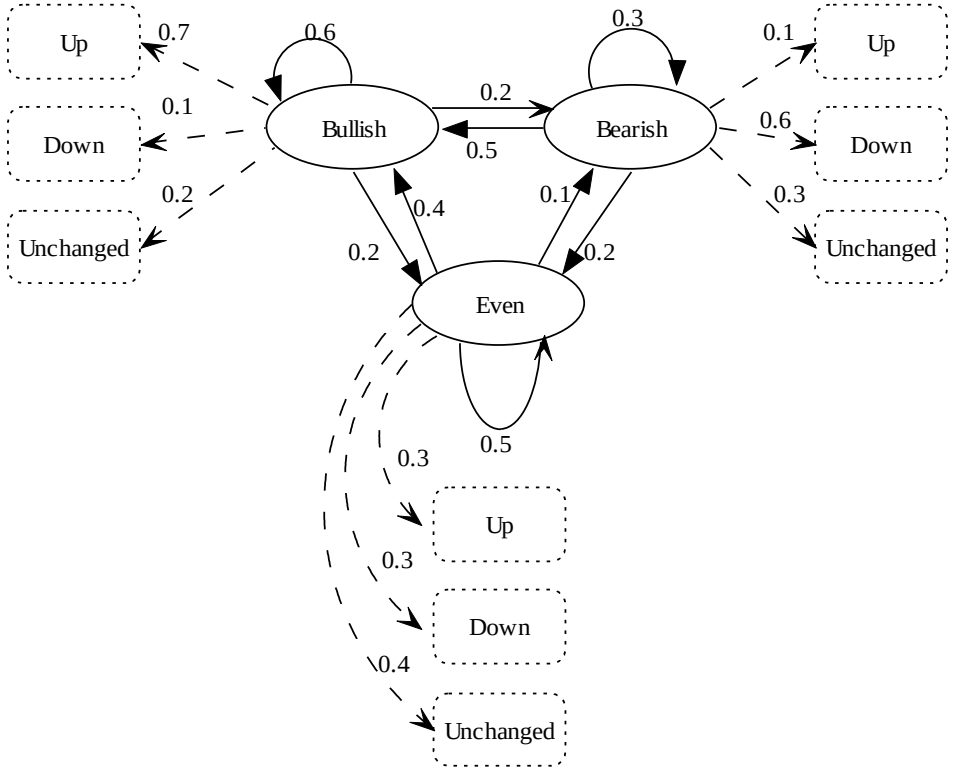

Fig. 3.2 – State Diagram representation of a First-Order Hidden Markov Model

The key point to note here is that we can only predict the hidden state by looking at the observable state; we cannot determine it completely. In the earlier simple case of Markov Process (where no state is hidden), up corresponded to bullish market and down to bearish market – there was no ambiguity. But, in this case, a definite probability always associates each hidden state with every observable state. As a result, a given observation sequence, let's say up-down-down-unchanged does not directly correspond to bullish-bearish-bearish-even. We cannot exactly determine what sequence of hidden states produced the observation sequence up-down-down, however, we can estimate as to which possible sequence of states is most likely to produce the given observed symbols/sequence.

3.2.2 HMM Parameters

The Hidden Markov model is represented as λ = (A, B, π) as defined by its A, B and π matrices. These three matrices are explained below.

1. **State Transition Probability Matrix (A):**

   This matrix contains values of the form a_{ij} where i is the source state and j is the destination state. Thus, a_{ij} corresponds to the probability of transition from state i to j. In other words, a_{ij} represents the probability of current state being state j (system being in state j at time t) provided the preceding state was i.

   Based on the nature of allowed transitions, there can be several types of HMMs.

   i. An HMM that allows transitions from every state to every other state of the HMM, i.e. all the coefficients of A are non-zero positive, is known as an ergodic HMM.

   ii. An HMM is said to be a left-to-right HMM when the entries of A satisfy the property $a_{ij} = 0$ for $j < i$, i.e., if the process can transit only to states whose indices are higher than the present state, the HMM is referred to as a left-to-right model. Essentially this means that the process is not allowed to go back to previous states. Thus, the above example would be a left-to-right HMM if one can only move from S_1 to S_1, S2 and S_3, from S_2 to S_2 and S_3, and from S_3 to S_3. In other words $a_{21} = a_{31} = a_{32} = 0$.

   The dimensions of this matrix are M X M.

2. **Observation symbol probability matrix (B):** The entries of this matrix $b_i(x)$ represent the probability of observing the symbol x while in state i. In the literature on HMMs, a re-

estimation procedure such as the Baum–Welch algorithm has been formulated [14] to estimate these values. The dimensions of this matrix are N X M.

3. **Initial state distribution/ Initial Probability Matrix (π):** This vector is given by $\pi = \{\pi 1, \pi 2, \ldots \pi N\}$, where π_i represents the probability of being in state *i* at the start of the observation sequence, or, the probability of the process being in the *ith* state when modelling of observations starts. The dimension of this matrix is 1 X N.

3.2.3 Notation

The following common notation is employed in the use of HMMs

- T: Temporal length / Length of observation sequence
- N: Number of hidden states in the HMM
- M: Number of possible observable states/observation symbols
- S: $\{S1, S2, \ldots, SN\}$ = Set of distinct states of the Markov process
- V: $\{v1, v2, \ldots, vM\}$ = Set of all possible observations
- A: State transition probability matrix = $\{ a_{i,j} \}$ for i < N, j < N.
- B: Observation symbol probability matrix = $\{ b_{i,j} \}$ for i < N, j < M.
- π: Initial probability matrix = $\{ \pi_i \}$, i < N = $\{\pi_1, \pi_2 , \pi_3 , \ldots, \pi_N \}$
- O: Observation sequence = $\{ O_i \}$, i < N = $\{O_1, O_2 , O_3 , \ldots, O_N \}$

Thus, for the above example, the parameters would be as follows

T = 4, N = 3, M = 3

S = $\{S1, S2, \ldots, SN\}$ (S1 = Bullish, S2 = Bearish, S3 = Even)

V = $\{v1, v2, \ldots, vM\}$ (V1 = Up, V2 = Down, V3 = Unchanged)

O = {up, down, down, unchanged} or {V1, V2, V2, V3}

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.5 & 0.3 & 0.2 \\ 0.4 & 0.1 & 0.5 \end{bmatrix} \quad , \quad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} 0.7 & 0.1 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$$

The sum of any particular row in both A and B is 1, as expected.

π = {0.33, 0.33, 0.33}

For the values of initial probability, we have assumed that it is equi-probable for the market to be Bullish (S1), Bearish (S2), or Even (S3) on the starting day. This matrix should be given in the process or calculated from past observations. Otherwise, equi-probability can be assumed for a stochastic process.

3.2.4 The Three Fundamental Problems of HMMs

HMMs can be used to address three fundamental types of problems. These problems are briefly described below along with the algorithms used to solve them. Out of these, problem types 2 and 3 are encountered in this work.

1. *Evaluation Problem:* Given the HMM model λ = (A, B, π) and an observation sequence O, find the likelihood that O is generated by the given HMM λ, i.e. P (O|λ). This is crucial for the recognition step and forward algorithm can be used to address it, but to make the process computationally faster, Viterbi algorithm is used instead.
2. *Decoding Problem:* Given the HMM model λ = (A, B, π) and an observation sequence O, find the most probable path, i.e. the sequence of hidden states that is most likely to have generated the given observation sequence. Viterbi decoding algorithm (Sec 2.3.2) is used to solve this.
3. Given an observation sequence O, find the HMM λ = (A, B, π) that maximizes P (O|λ), i.e. the probability of O given λ. (best fitting of observed data). In this case, the parameters of the HMM are iteratively updated so as to make it sufficiently representative of the observation sequence. During this fitting step, Baum-Welch algorithm (Sec 2.3.3) is used for the re-estimation of parameters A, B and π.

3.3 HMM algorithms

This section briefly lists the two HMM algorithms that have been used in this work for training and recognition.

3.3.1 Viterbi decoding:

It is used to find the most probable path or Viterbi path of the process, i.e. the sequence of hidden states that has the maximum likelihood of having generated the given observation sequence.

Suppose the HMM has state space S = {S_1, S_2, …, S_N} and initial probabilities π = {π_1, π_2 , π_3 ,…, π_N }. Let the transition probability matrix be A = {$a_{i,j}$} for i, j < N. If the observation sequence

is $\{O_1, O_2, \ldots, O_T\}$ (T = temporal length of sequence = number of frames), then the most likely path $X_1, X_2, \ldots, X_T$ that produces the given observation sequence O is given by:

$$V_{1,k} = P(O_t \mid k).\pi_k$$

(3.3.1.1)

$$V_{t,k} = \max_{x \in S}(P(O_t \mid k).a_{x,k}.V_{t-1,x})$$

(3.3.1.2)

Where $V_{t,k}$ denotes the maximum probability (out of all possible state sequences) of generation of the observed sequence. Thus, we can retrieve the 'maximum probable path' by assigning back pointers to the values of x used in Eq. 3.3.1.2 and subsequent ranking.

3.3.2 Baum-Welch algorithm:

This algorithm is used to re-estimate the unknown parameters A, B and π after each iteration during the training process so as to refine the HMM model iteratively to fit the observed sequence. Given a particular observation sequence, it uses expectation maximization in order to find the maximum likelihood estimate of HMM parameters.

If X_t is the state at time t, then

$$A = \{a_{i,j}\} = P(X_t = j \mid X_{t-1} = i)$$

(3.3.2.1)

Let the initial probability and observation sequence be defined in the same as in Sec. 3.3.1. Now, the Baum Welch algorithm is used to maximize $P(O \mid \lambda)$ in the following manner.

A set of initial values are assigned to the parameters (A, B, π) at the beginning of the process. The details of initialization for this particular work are given in Sec. 3.2.

- Forward Procedure: $\alpha_i(t) = P(O_1, O_2, ..., O_t \mid X_t = i)$ is the probability of having observations $O_1, O_2, ..., O_t$ from time 1 to t while being in the ith state at time t. This is computed recursively as $\alpha_j(t+1) = b_j(O_{t+1}) \sum_{i=1}^{N} \alpha_i(t)\alpha_{ij}$.
- Backward Procedure: $\beta_i(t) = P(O_{t+1}, O_{t+2}, ..., O_T \mid X_t = i)$ is the probability of having the observations of partially ending sequence as $O_{t+1}, O_{t+2}, ..., O_T$ from time t+1 to T when the system is in the ith state at time t. $\beta_i(t)$ is computed as $\beta_i(t) = \sum_{j=1}^{N} \beta_j(t+1)\alpha_{ij} b_j(O_{t+1})$

- Update Procedure: The temporary variables can be calculated using Bayes algorithm

$$\text{as } \gamma_i(t) = P(X_t = i \mid O, \lambda) = \frac{\alpha_i(t)\beta_i(t)}{\sum_{j=1}^{N} \alpha_j(t)\beta_j(t)}$$

(3.3.2.2)

where $\gamma_i(t)$ is the probability of being in state i at time t given the HMM λ and the observed sequence O.

$$\text{Now, } \xi_{i,j}(t) = P(X_t = i, X_{t-1} = j \mid O, \lambda)$$

(3.3.2.3)

gives the probability of the system successively being in states i and j at time instants t and t+1 respectively, given the HMM λ and the observed sequence O.

The HMM parameters can now be updated by using the following equations

Initial probability: $\pi_i^* = \gamma_i(1)$ (3.3.2.4)

Transition probability: $\alpha_{ij}^* = \frac{\sum_{t=1}^{T-1} \xi_{ij}(t)}{\sum_{t=1}^{T-1} \gamma_i(t)}$

(3.3.2.5)

In this work, some case-specific equations as described in Sec. 3.2 have been used for the updating of parameters.

3.4 Training and Recognition Procedure using HMM

In the direct approach, initial estimates of ε and HMM parameters are obtained from the observed sequence. Subsequently, these are iteratively updated using Expectation Minimization (EM). The entire procedure can be summarized into three broad steps as follows -

Step 1: Initialization

For the initial estimates, the gait sequence is segmented into several gait cycles and subsequently each gait cycle is divided into N temporally adjacent segments that are nearly equal in size. An initial estimate of the ith exemplar e_i is obtained by taking the mean of the feature vectors of all

the frames present in the ith cluster. For the transition matrix A, an initial estimate is obtained by assuming that from every state, the gait process can only move to the next state or remain in the same state (since walking is a quasi-periodic process). The only exception is a possible transition from the last state S5 to the first state S1 Thus, the process can move as illustrated in Fig. 3.3.

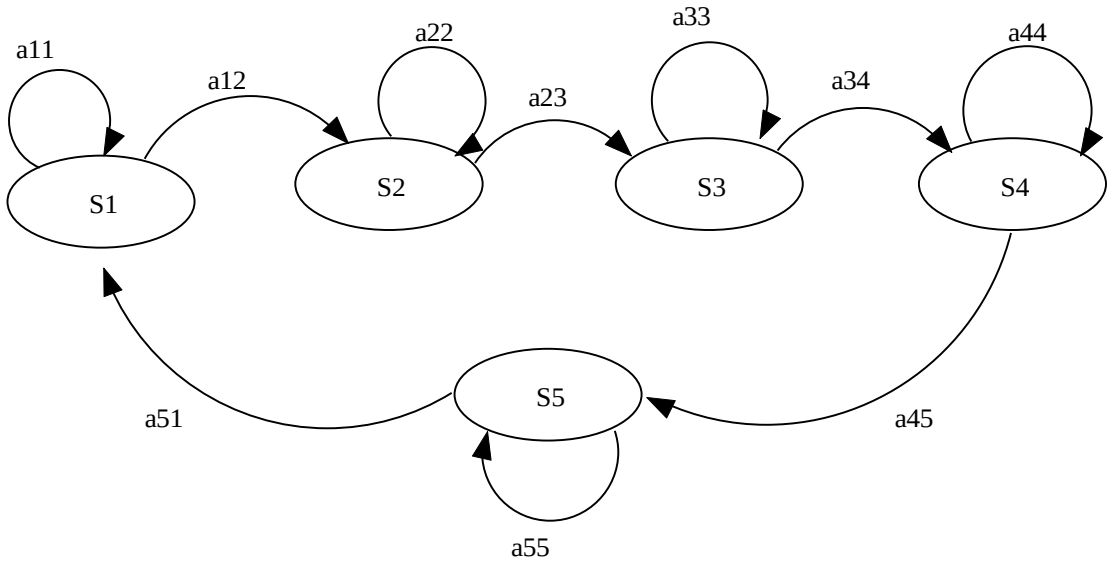

Fig 3.3 – State Transition Diagram of a characteristic gait sequence

Now it is assumed that no transitions except the ones above are allowed. Therefore the probabilities associated with all the transitions except the aforementioned are set to zero. Also, it is assumed that it is equally probable for the process to remain in a state or to move to the next state. For N = 5, this gives a transition probability matrix as shown below.

$$A = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0 & 0 & 0.5 \end{bmatrix}$$

For the initial estimates of the initial probability matrix, it is assumed that it is equally probable for the process to start from any one state, thus each state is assigned the initial probability value of 1/N. For N= 5, this gives an initial probability matrix as follows

$$\pi = \{0.2, 0.2, 0.2, 0.2, 0.2\}$$

The Observation symbol probability matrix B is defined in terms of an exponential function of distance measure (Inner Product Distance (IPD)) between the frame feature vectors and the exemplars, and is given by

$$b_i(f(t)) = \alpha\delta_i e^{-\delta_i D(f(t),e_i)}$$

(3.4.1)

Where $f(t)$ is the frame at time t, i = 1 : N, and $D(f(t),e_i)$ represents the inner product distance between f(t) and the ith exemplar. α < 1 is a constant and δ_i is given by

$$\delta_i = \frac{N_i}{\sum_{f(t)\in C_i} D(f(t),e_i)}$$

(3.4.2)

where N_i is the number of gait frames in the ith cluster. This completes the initialization step.

Step 2: Training

For a sample observed sequence, Viterbi algorithm (refer Sec. 3.3.2) is used to obtain the most probable path of hidden states, say $Q = \{q(1), q(2), ..., q(NF)\}$, where NF denotes the number of frames in the gait cycle. Subsequently, the exemplar vectors are updated as follows:

From the most probable path, we can get a set of frames belonging to each cluster. Let T_n represent the set of time instants where f(t) belongs to the nth exemplar. For example, if there are 10 frames and the most probable path sequence is $\{S_1, S_1, S_1, S_2, S_3, S_3, S_4, S_4, S_5, S_1,\}$, then the set of frames representing cluster 1 or exemplar 1 will be $\{f(1), f(2), f(3), f(10)\}$. Thus $T_1 = \{1,2,3,10\}$.

The exemplars are updated such that the probability of a frame belonging to the estimated exemplar (as determined from most probable path) is maximized, or the distance between the frame and the exemplar is minimized.

$$e_n^{i+1} = \arg_e \max \prod_{t\in T_n} P(f(t)\,|\,e) \qquad or \qquad e_n^{i+1} = \arg_e \min \prod_{t\in T_n} D(f(t)\,|\,e)$$

(3.4.3)

where e_n^{i+1} represents the nth exemplar in the (i+1)th iteration. The update is accomplished by

$$e_n^{i+1} = \sum_{t\in T_n} \tilde{f}(t)$$

(3.4.4)

where $\tilde{f}(t)$ represents the normalized frames belonging to $t \in T_n$. This simply means that the new nth exemplar is obtained by averaging the frames contained in the frame set T_n.

From these new exemplars, the new observation symbol probability matrix B can be calculated by using Eq. 3.4.1 and 3.4.2. Thereafter, parameters A and π can be iteratively updated using the Baum-Welch algorithm (refer Sec. 3.3.2).

Step 3: Classification/Recognition

In this step, the HMM parameters as well as exemplars of all the persons in the database are available, along with a probe gait sequence. The objective is to reveal the identity of the subject in the probe gait sequence by finding the HMM that has the maximum probability of generating that particular sequence.

Initialization for the parameters A and π remains the same as in the training step while initial value for observation symbol probability matrix B is obtained using the exemplars of each person in the gait database. For each person in the database, we already have the HMM parameters and exemplar vectors, and Viterbi algorithm is used to compute the likelihood or probability that the probe sequence was produced by the HMM of that person. A ranking is carried out using these similarity scores.

Fig. 3.4.a and 3.4.b illustrate the training procedure and recognition procedure respectively in the form of flowcharts. Here, P is the total number of persons in the gait database.

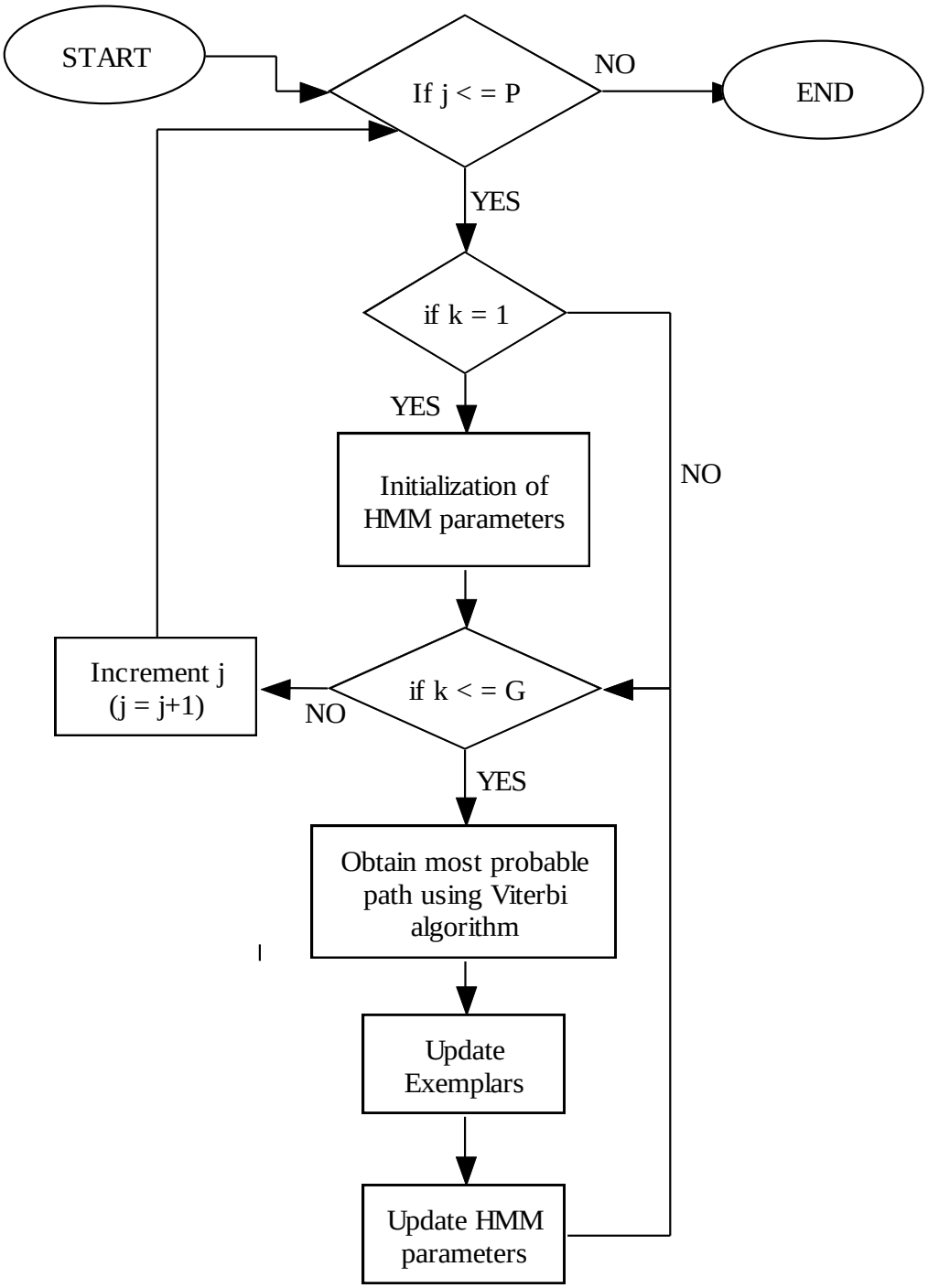

Fig. 3.4.a – Flowchart representation of HMM-based training procedure

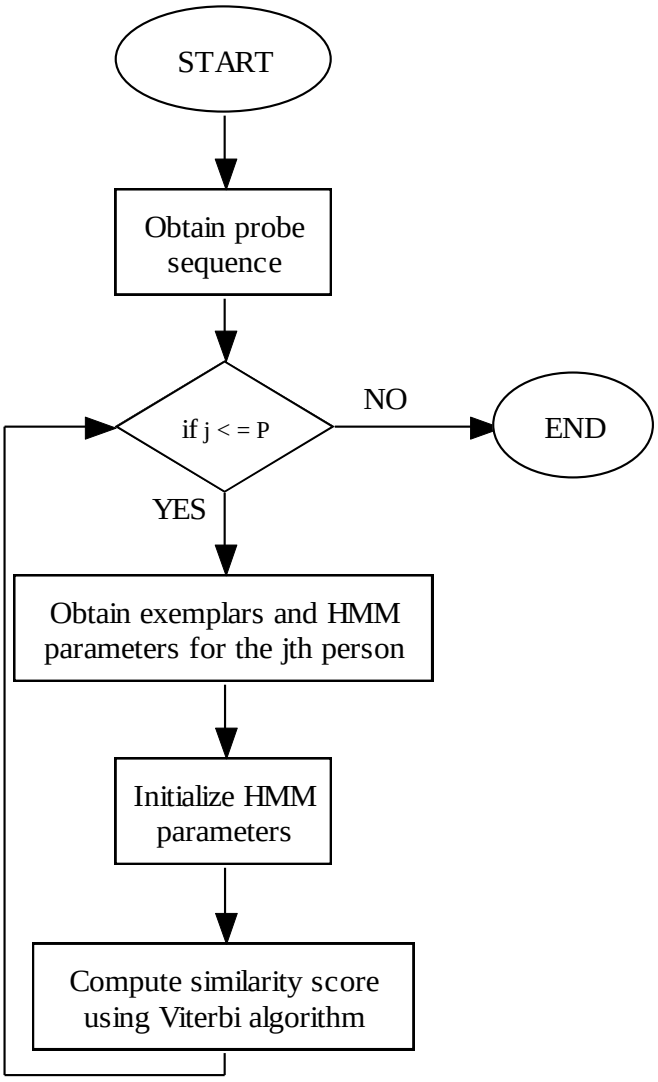

Fig. 3.4.b – Flowchart representation of recognition part of methodology

3.5 Results and Discussion

Table 3.1 shows the results of HMM-based recognition on the gait data of 118 persons taken from the CASIA Dataset-B. While [2] uses the CASIA dataset, [1] uses a different dataset.

Table 3.1 Overall Experimental results using i) direct and indirect approach and ii) with and without FDEI reconstruction

Method	Without FDEI reconstruction	With FDEI reconstruction
Direct Approach (wavelet)	80.93 %	86.44 %
Indirect Approach (FED)	72.43 %	71.39 %

There are 4 normal gait sequences of each person in the CASIA Dataset (nm_01, nm_02, nm_03 and nm_04). Out of these, three are used as training sets and the remaining 1 as the test/probe sequence. We use 4-fold cross-validation to ensure the reliability of the process. The results with and without FDEI reconstruction are tabulated below.

Table 3.2 – Experimental results using 4-fold cross validation in CASIA Dataset-B using i) direct and indirect approach and ii) with and without FDEI reconstruction

With FDEI reconstruction				
	Probe Set:nm_1	Probe Set:nm_2	Probe Set:nm_3	Probe Set:nm_4
HMM with wavelet feature	90.68 % (107/118)	83.05 % (98/118)	83.9 % (99/118)	88.14 % (104/118)
HMM with FED vectors	75.42 % (89/118)	69.49 % (82/118)	66.95 % (79/118)	73.73 % (87/118)
Without FDEI reconstruction				
HMM with wavelet feature	88.13 % (104/118)	77.12 % (91/118)	73.73 % (87/118)	84.75 % (100/118)
HMM with FED vectors	77.12 % (91/118)	68.64 % (81/118)	66.95 % (79/118)	72.88 % (86/118)

The recognition performance depends also on the number of exemplars N, which is equal to the number of hidden states. Comparing the results using Probe Set nm_01, we obtain the optimal results at N = 5.

Table 3.3 – Variation of recognition performance with change in N

	With FDEI reconstruction		Without FDEI reconstruction	
N	Direct Approach	Indirect Approach	Direct Approach	Indirect Approach
3	74.5	57.7	63.2	56.6
4	81.2	66.3	76.4	67.2
5	**86.4**	**71.4**	**80.9**	**72.4**
6	83.2	70.1	80.9	70.5

The result deteriorates significantly when N < 4. For N = 5 and N = 6, the results are relatively close. Also, as N changes, the variation in performance is more in the indirect approach as the N directly determines the number of elements in the FED vector.

Recognition rates also vary as the viewing angle changes. The viewing angle remains fixed for a given run. We test the procedure only for norm_01 set with six different viewing angles. The optimal results are obtained at an angle of 0 degree.

Table 3.4 – Variation of recognition performance with change in viewing angle (norm_01)

Viewing Angle (degrees)	With FDEI reconstruction		Without FDEI reconstruction	
	Direct Approach	Indirect Approach	Direct Approach	Indirect Approach
0	**86.4**	**71.2**	**80.5**	**72.4**
36	84.7	70.3	81.4	69.5
54	83.0	68.6	78.8	66.9
90	83.9	68.6	77.1	65.2
126	81.4	66.9	76.3	65.2
180	85.6	71.2	81.4	70.3

Surprisingly, 90 degree has a lower recognition performance compared to 0 degree even when the proportion of dynamic information is maximum. This is probably due to greater probability of imperfections in 90 degree images due to increased viewing area.

Cumulative Match Characteristic (CMC): A Cumulative Match Characteristic (CMC) curve [1] [2] is used to plot the probability of identification of a class/subject against the 1: P class list (assuming P number of classes). It gives the likelihood of a given class/user appearing in candidate/class lists of size 'n', where n = 1 : P. The faster the CMC curve approaches the score of 100 (meaning the particular class/subject always appears in the first n-ranked classes), the better the matching algorithm.

Since we use 4-fold Cross Validation, we compute the Rank for all the four cycles, such that the total number of cases is 472. Table 3.3 shows the number of frames identified correctly up to the nth rank $(n \in 1,5,10,...,30,40)$. Although we use the comprehensive ranking for plotting the CMC curves, for tabulation we only include ranks up to 40 for practical reasons. The maximum gradient is found in the initial 20 ranks after which the rate of change of CMC curve gradually slows down.

Table 3.5 – n-Rank Cumulative Match Scores using Direct and Indirect approaches

Rank	Direct Approach	Indirect approach
1	408 (86.44)	337 (71.39)
5	449 (95.13)	408 (84.11)
10	462 (97.88)	432 (91.53)
15	467 (98.94)	461 (97.67)
20	471 (99.79)	469 (99.36)
25	472 (100)	469 (99.36)
30	472 (100)	469 (99.36)
35	472 (100)	471 (99.79)
40	472 (100)	472 (100)

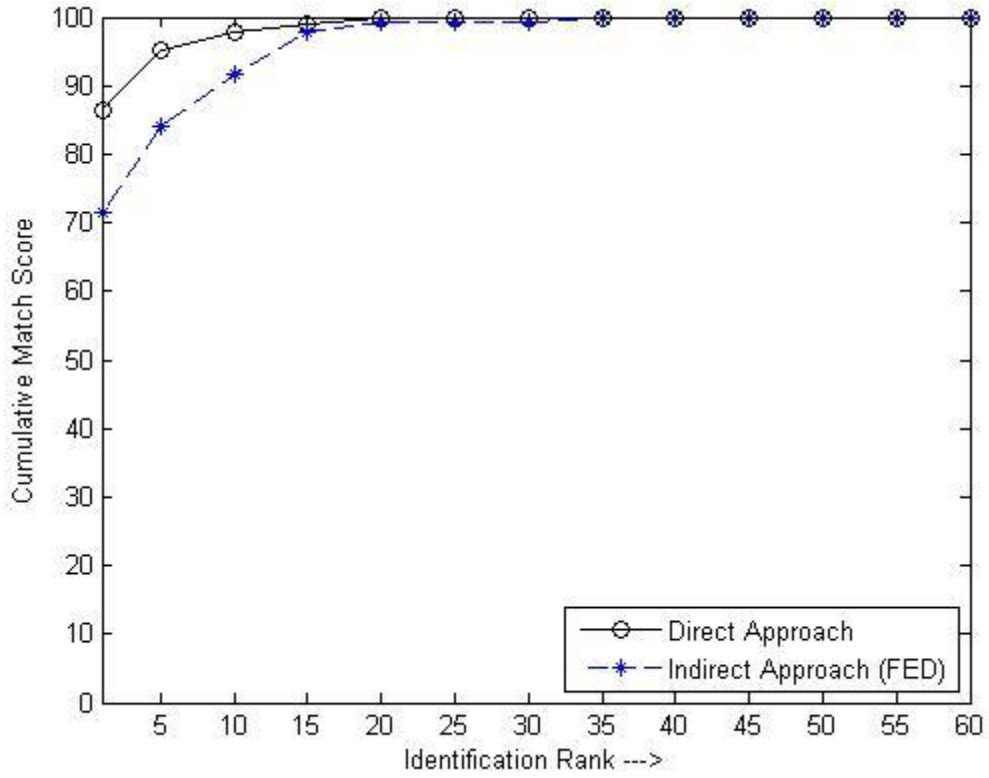

Fig. 3.5 – Cumulative Match Characteristic (CMC) curve of the experimental results

The results obtained when FDEI reconstruction is applied is better compared to the results obtained using the original binary silhouettes, both for direct and indirect approaches, and conforms to the observations made in [2].

As expected, the results using the direct approach are better, owing to a basic trade-off between complexity and efficiency. The direct approach is computationally more extensive with a better overall performance, while the indirect approach tends to be computationally less intensive at the cost of recognition performance. Also, as discussed in [1], the direct approach is less vulnerable to noise and distortion compared to the indirect approach.

In the CMC curve also, the curve representing the direct method approaches 100% fairly quickly compared to the indirect method.

Surprisingly, the results obtained by using the FED vector show that in this case, the overall performance is actually slightly lesser in the case of FDEI reconstruction as compared to the use of direct binary silhouettes.

3.6 Conclusion

This chapter describes the training and recognition steps using Hidden Markov Models (HMMs) as part of the overall gait identification framework. Recognition is carried out using both the direct and indirect approaches of feature extraction and also including/excluding computation of Frame Difference Energy Image. Results show an increase in recognition performance with inclusion of the FDEI reconstruction step and with the use of wavelet approximation coefficients as direct features. Recognition is also seen to vary across change in parameters like camera view angle and N, with the best figures being obtained at a viewing angle of 0 degree (subject walking towards camera) and N = 5. The maximum identification accuracy reached by the system is 86.44 %, when direct approach is combined with FDEI reconstruction.

CHAPTER 4

CONCLUSIONS AND FUTURE WORK

4.1 Conclusions

This work explores a gait recognition method with binary silhouette-based input images and Hidden Markov Model (HMM)-based classification. The performance of the recognition method depends significantly on the quality of the extracted binary silhouettes. A fuzzy correlogram based method is employed for background subtraction and Frame Difference Energy Image (FDEI) reconstruction is performed to make the recognition method robust to partial incompleteness of silhouettes. Feature extraction is performed in two ways – direct and indirect. The direct method uses extracted features directly for classification while the indirect method maps the higher-dimensional features into a lower-dimensional space by means of a Frame-to-Exemplar-Distance (FED) vector. The FED uses the distance measure between pre-determined exemplars and the feature vectors of the current frame as an identification criterion. The extracted features are fed to the HMM-based classifier which models human gait as a First-Order Markov Process and uses Viterbi decoding and Baum-Welch algorithm to compute similarity scores and carry out identification. This work achieves an overall sensitivity of 86.44 % using the direct approach and 71.39 % using the indirect approach. Without the application of FDEI reconstruction, the procedure achieves a lesser overall sensitivity of 80.93 % using the direct approach and 72.43 % using the indirect approach, as per expectations.

As expected, the results using the direct approach are better, owing to a basic trade-off between complexity and efficiency. The direct approach is computationally more extensive with a better overall performance, while the indirect approach tends to be computationally less intensive at the cost of recognition performance. Also, as discussed in [1], the direct approach is less vulnerable to noise and distortion compared to the indirect approach.

A CMC curve is plotted to measure the reliability of the process. In the CMC plot, the curve representing the direct method approaches 100% fairly quickly compared to the indirect method.

Surprisingly, the results obtained by using the FED vector show that in this case, the overall performance is actually slightly lesser in the case of FDEI reconstruction as compared to the use of direct binary silhouettes.

It is also seen that the recognition performance depends to a certain extent on the specific gait sequences. The algorithms used in this work produce relatively better results in the case of norm_01 and norm_04 sets of the CASIA Dataset-B compared to norm_02 and norm_03 sets. Variation in recognition performance is also observed with change in characteristic parameters

like viewing angle and N and the best results are obtained when the path of subject parallel to camera axis (viewing angle of 0 degree) and at N = 5.

4.2 Future Work

The major objective of this work is to implement an efficient algorithm for the reliable identification of humans using their gait data. The experimental results provide a preliminary platform for satisfactory gait identification, but there exists a huge scope for future work in this area especially regarding issues of real-world implementation. Some possible additional works that can be done are –

i. The classification procedure can be extended to other classifier models such as Gaussian Mixture Models (GMMs).
ii. A database having a larger number of subjects could be employed to confirm the reliability of the algorithm and make it more robust.
iii. Scope for real-time implementation should be investigated, employing features and processes that are computationally less intensive. This can be useful for biometric attendance systems.
iv. Occlusions are a major concern in surveillance sequences involving more than one person, especially in public places. Algorithms should be made insensitive to occlusions and case-specific fluctuations like clothing and carrying conditions.

References

[1] A. Kale, A. Sundaresan, A.N. Rajagopalan, “Identification of Humans using Gait,” *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1163-1173, Sept. 2004.

[2] C. Chen, J. Liang, H. Zhao, H. Hu and J. Tian, “Frame Difference Energy Image for gait recognition with incomplete silhouettes,” *Pattern Recognition Letters*, vol. 30, pp. 977-984, May 2009.

[3] J. Xue, D. Ming, W. Song, “Infrared gait recognition based on wavelet transform and support vector machine,” *Pattern Recognition*, vol. 43, pp. 2904-2910, Mar. 2010.

[4] A. Kale, N. Cuntoor and R. Chelappa, “A Framework for Activity-specific Human identification,” Univ. of Maryland at College Park.

[5] C. Pojala, S. Sengupta, “Detection of moving objects using Fuzzy Correlogram based, background subtraction”, *Int. Conf. on Signal and Image Processing Application (ICSIPA)*, 2011

[6] L. Lee and W. E. L. Grimson, “Gait analysis for recognition and classification,” in Proc. *IEEE Conf. Face and Gesture Recognition*, pp. 155–161, 2002.

[7] D. Cunado, J. M. Nash, M. S. Nixon, and J. N. Carter, “Gait extraction and description by evidence-gathering,” In Proc. *Int. Conf. Audio and Video Based Biometric Person Authentication*, pp. 43–48, 1995.

[8] P. S. Huang, C. J. Harris, and M. S. Nixon, “Recognizing humans by gait via parametric canonical space,” Artif. Intell. Eng., vol. 13, no. 4, pp. 359–366, Oct. 1999.

[9] J. Little and J. Boyd, “Recognizing people by their gait: The shape of motion,” Videre, vol. 1, no. 2, pp. 1–32, 1998.

[10] J. Han and B. Bhanu, "Individual recognition using gait energy image," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, pp. 316-322, 2006.

[11] W. Liang, T. Tieniu, N. Huazhong, and H. Weiming, "Silhouette analysis-based gait recognition for human identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1505-1518, 2003.

[12] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanID gait challenge problem: data sets, performance, and analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 162-177, 2005.

[13] L. Zongyi and S. Sarkar, "Improved gait recognition by gait dynamics normalization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 863-876, 2006.

[14] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, pp. 257–285, 1989.

[15] Han, J., Bhanu, B., "Individual recognition using Gait Energy Image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28 (2), pp. 316–322, 2006.

[16] Elgammal, A., Harwood, D., Davis, L., "Non-parametric model for background subtraction," *Sixth European Conf. on Computer Vision*, Dublin, Ireland, 2000.

[17] Friedman, N., Russell, S,"Image segmentation in video sequences: A probabilistic approach," *13th Conf. on Uncertainty in Simulating Intelligence.* Rhode Island, USA, pp. 175-181.

[18] Tian, T.Y., Shah, M., "Motion estimation and segmentation," *Machine Vision Appl. (Springer)*, pp. 32–42, 1996.

[19] Yacob, Y., Black, M.J., "Parameterized modeling and recognition of activities," *6th Internat. Conf. on Computer Vision. India*, pp. 120–127, 1999.

[20] Liu, Z., Malave, L., Sarkar, S., "Studies on silhouette quality and gait recognition," *IEEE Internat. Conf. on Computer Vision and Pattern Recognition*, pp. 704–711, 2004.

[21] Liu, Z., Sarkar, S., "Effect of silhouette quality and hard problems in gait recognition," *IEEE Trans. Systems Man and Cybernetics.* Part B: Cybernetics, vol. 35 (2), pp. 170–183, 2005.

[22] Yu, S., Tan, D., Huang, K., Tan, T., "Reducing the effect of noise on human contour in gait recognition," *Internat. Conf. on Biometrics*, pp. 338–346, 2007.

[23] Liu, J., Zheng, N., "Gait History Image: A novel temporal template for gait recognition," *IEEE Internat. Conf. on Multimedia and Expo*, pp. 663–666, 2007.

[24] Ma, Q., Wang, S., Nie, D., Qiu, J., "Recognizing humans based on Gait Moment Image," *8th ACIS Internat. Conf. on SNPD*, pp. 606–610, 2007.

[25] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence* vol. 23, pp. 257-267, 2001.

[26] J. Liu, N. Zheng, and L. Xiong, "Silhouette quality quantification for gait sequence analysis and recognition," *Signal Processing,* vol. 89, pp. 1417-1427, 2009.

[27] Z. Xue, D. Ming, W. Song, B. Wan, and S. Jin, "Infrared gait recognition based on wavelet transform and support vector machine," *Pattern Recognition,* vol. 43, pp. 2904-2910, 2010.

[28] A. Kale, A. N. Rajagopalan, N. Cuntoor, and V. Kruger, "Gait-based recognition of humans using continuous HMMs," in *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 336-341, 2002.

[29] W. Liang, T. Tieniu, N. Huazhong, and H. Weiming, "Silhouette analysis-based gait recognition for human identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1505-1518, 2003.

[30] F. Dadashi, B. N. Araabi, and H. Soltanian-Zadeh, "Gait Recognition Using Wavelet Packet Silhouette Representation and Transductive Support Vector Machines," in *2nd International Congress on Image and Signal Processing*, pp. 1-5, 2009.

[31] N. V. Boulgouris, K. N. Plataniotis, and D. Hatzinakos, "An angular transform of gait sequences for gait assisted recognition," in *IEEE International Conference on Image Processing*, vol.2. 32326, pp. 857-860, 2004.

[32] W. Liang, T. Tieniu, H. Weiming, and N. Huazhong, "Automatic gait recognition based on statistical shape analysis," *IEEE Transactions on Image Processing*, vol. 12, pp. 1120-1131, 2003.

[33] N. V. Boulgouris and Z. X. Chi, "Gait Recognition Using Radon Transform and Linear Discriminant Analysis," *IEEE Transactions on Image Processing*, vol. 16, pp. 731-740, 2007.

[34] L. R. Rabiner, "A tutorial on hidden Markov models andselected applications in speech recognition," *Proceedings of\the IEEE*, vol. 77, pp. 257-286, 1989.

[35] M. A. Moni and A. B. M. S. Ali, "HMM based hand gesture recognition: A review on techniques and approaches," in *2nd IEEE International Conference on Computer Science and Information Technology*, pp. 433-437, 2009.

[36] S. Aravind, R. Amit, and C. Rama, "A hidden Markov model based framework for recognition of humans from gait sequences," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 3, pp. II-93-6, 2003.

[37] H. Qiang and C. Debrunner, "Individual recognition from periodic activity using hidden Markov models," in *Proceedings of the Workshop on Human Motion*, pp. 47-52, 2000.

[38] Yin Pei, I. Essa, T. Starner, and J. M. Rehg, "Discriminative feature selection for hidden Markov models using Segmental Boosting," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2001-2004, 2008.

[39] C. Changhong, L. Jimin, Z. Heng, H. Haihong, and T. Jie, "Factorial HMM and Parallel HMM for Gait Recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews,* vol. 39, pp. 114-123, 2009

[40] C. Ming-Hsu, H. Meng-Fen, and H. Chung-Lin, "Gait Analysis for Human Identification through Manifold Learning and HMM," in *IEEE International Symposium on Circuits and Systems*, pp. 969-972, 2007.

[41] L. Zongyi and S. Sarkar, "Effect of silhouette quality on hard problems in gait recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 35, pp. 170-183, 2005.

[42] L. Zongyi, L. Malave, and S. Sarkar, "Studies on silhouette quality and gait recognition," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004, vol. 2, pp. II-704-II-711