

University of Nevada, Reno

# **Feature Selection Using Genetic Algorithms for Human Gait Recognition**

A thesis submitted in partial fulfillment of  
the requirements for the degree of Master  
of Science in Computer Science and  
Engineering

By

Faezeh Tafazzoli

Professor George Bebis/Thesis Advisor

May 2012



University of Nevada, Reno  
Statewide • Worldwide

THE GRADUATE SCHOOL

We recommend that the thesis  
prepared under our supervision by

**FAEZEH TAFAZZOLI**

entitled

**Feature Selection Using Genetic Algorithms  
for Human Gait Recognition**

be accepted in partial fulfillment of the  
requirements for the degree of

**MASTER OF SCIENCE**

George Bebis, Ph.D., Advisor

Mircea Nicolescu, Ph.D., Committee Member

Mark A. Pinsky, Ph.D., Graduate School Representative

Marsha H. Read, Ph. D., Dean, Graduate School

May, 2012

## ABSTRACT

Many research studies have demonstrated that gait can serve as a useful biometric feature for human identification at a distance. Here we manifest the importance of feature selection in gait recognition systems. Feature selection is an important factor which impacts the classification accuracy. This goal is achieved by discarding irrelevant and redundant information which affects both the classifier's performance and system's efficiency. Traditional gait recognition systems have mostly been evaluated without considering the most relevant features. In this study, we are going to investigate the use of Genetic Algorithm (GA) for selecting an optimal subset of features for a model-free gait recognition approach without degrading the classification accuracy. First, features are extracted using Kernel Principal Component Analysis (KPCA) on four spatiotemporal projections of silhouettes. Then, GAs are applied to choose a subset of Eigen-vectors that represent a subject's identity. Our experimental results, conducted on Georgia Tech (GT) database, indicate considerable performance improvements.

**Keywords:** Human identification, Gait, Genetic Algorithms, Feature selection, Principal Component Analysis.

## **ACKNOWLEDGMENTS**

It is with great pleasure that I reflect on my time at UNR about all the people who made this thesis possible. First and foremost, I would like to thank my advisor, Professor Bebis, for his patient, thoughtful guidance and support.

I would also like to express my deep appreciation to the rest of my thesis committee: Professor Mircea Nicolescu and Professor Mark Pinsky for their direction and insightful comments and invaluable advice.

I am especially thankful to Professor Sushil Louis for his dedication and sharing knowledge which caused me to take the first steps of this project in his valuable course.

I would like to convey my gratitude to all the participant teachers, students and coordinators who formed part of this study.

# CONTENTS

<b>Abstract</b> .....	i
Acknowledgments.....	ii
Contents .....	iii
List of Figures.....	iv
List of Tables .....	v
<b>CHAPTER 1 Introduction</b> .....	<b>1</b>
<b>CHAPTER 2 Literature Review</b> .....	<b>5</b>
2.1 Model-Based Approaches .....	5
2.2 Model-Free Approaches.....	9
<b>CHAPTER 3 Feature Selection</b> .....	<b>17</b>
3.1 GA-Based Feature Selection .....	19
<b>CHAPTER 4 Method Overview</b> .....	<b>22</b>
4.1 Gait Pattern Extraction and Representation .....	23
4.1.1 Preprocessing.....	23
4.1.2 Gait Representation .....	28
4.2 Feature Extraction using Kernel PCA.....	30
4.3 Genetic Feature Subset Selection.....	32
4.3.1 Feature Selection Encoding.....	34
4.3.2 Fitness Evaluation .....	34
4.3.3 Initial Population .....	36
4.3.4 Selection .....	36
4.3.5 Operators .....	37
4.4 Gait Recognition.....	38
<b>CHAPTER 5 Experimental Results</b> .....	<b>41</b>
5.1 Dataset .....	41
5.2 Algorithm in Steps.....	42
<b>CHAPTER 6 Conclusion And Future Work</b> .....	<b>54</b>
<b>Bibliography</b> .....	<b>55</b>

## LIST OF FIGURES

<b>Fig. 1. 1</b> Marker Positions in Gait Analysis .....	2
<b>Fig. 2. 1</b> Distinction of (a) Static and (b) Dynamic Parameters.....	6
<b>Fig. 4. 1</b> Proposed Algorithm Framework .....	23
<b>Fig. 4. 2</b> Preprocessing Framework.....	24
<b>Fig. 4. 3</b> Constructed Graph for a Sample 3×3 Portion of the Image Plane [71] .....	26
<b>Fig. 4. 4</b> Global Temporal Accumulation [70].....	27
<b>Fig. 4. 5</b> Silhouette Representation [72].....	29
<b>Fig. 5. 1</b> Sample Sequences of GT Database. ....	42
<b>Fig. 5. 2</b> Output of Foreground Extraction and Moving Edges Detection. ....	43
<b>Fig. 5. 3</b> Bulk Motion of Silhouette for a Sample Sequence for Different Velocity Values.....	44
<b>Fig. 5. 4</b> Variation of Total Intensity based on Velocity in Silhouette Accumulation.....	45
<b>Fig. 5. 5</b> Periodic Pattern of Silhouette Width Variation for Two Sample Sequences. ....	46
<b>Fig. 5. 6</b> A Sample Silhouette and its Four Projections for a Selected Frame of Sequence side0001m03.....	46
<b>Fig. 5. 7</b> Spatio-temporal Gait Patterns estimated for Top, Bottom, Left and Right Projections for the Sequence side0001m03. ....	47
<b>Fig. 5. 8</b> Results of GA subset selection, population size: 800, generation size: 50. Pc: 0.6, Pm: 0.1, 7-fold cross validation, Comparing the results of different classifiers.....	49
<b>Fig. 5. 9</b> Results of GA subset selection, population size: 1500, generation size: 20. Pc: 0.6, Pm: 0.1, 7-fold cross validation, Comparing the results of different selection algorithms. ....	49
<b>Fig. 5. 10</b> Results of GA subset selection, population size: 1000, generation size: 100, Naive Bayes, elitism selection, 7-fold cross validation, Comparing the results of different mutation and recombination probability values. ....	50

## LIST OF TABLES

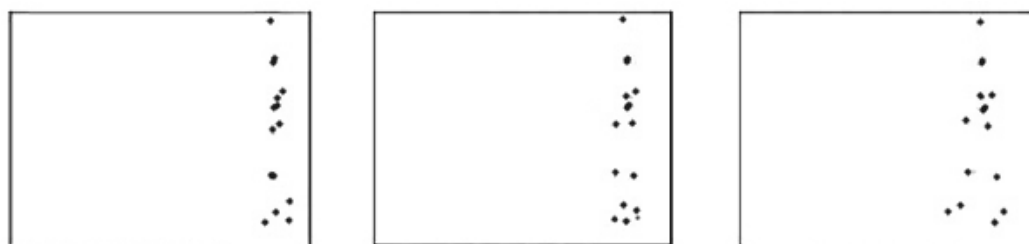
<b>Table 5.1</b>	Averaged Performance in all 6 Experiments before GA.....	51
<b>Table 5.2</b>	Overall Performance after GA.....	51
<b>Table 5.3</b>	Comparison of proposed algorithm with some other approaches.....	53

# CHAPTER 1

## INTRODUCTION

In today's modern societies, a reliable automatic means of human identification in many different situations, such as surveillance and access control is in great demand. Biometrics, an identification method based on human's physiological or behavioral features, has come to occupy an increasingly important role in identification, due to their universality and uniqueness [1]. The primary advantage of biometric identification over other methods of identification which rely on "something you possess" or "something that you know" is that biometrics cannot be misplaced or forgotten. For personal identification, many biometric-based authentication methods have emerged using a wide variety of cues, such as fingerprint, iris, face, palm print, handwriting, voice, gait, or a combination of these traits. Of these, gait identification, which is defined as a means of identifying individuals by the way they walk, is appealing from a surveillance standpoint due to its ability to ascertain identity at a distance, along with being noninvasive and non-perceivable [2, 3]. Gait also can be detected and measured at low resolution, and therefore it can be used in situations where face or iris information is not available in high enough resolution for recognition or human subject occupies too few image pixels. In comparison with other biometric features, also it is less likely to be obscured [4]. Furthermore, night vision capability is usually not possible with other biometrics. There is much evidence from psychophysical, medical and

biomechanical experiments that each individual has his own characteristic gait pattern [5]; moreover, there is ample anecdotal evidence about human's ability to identify acquaintances based only on their manner of walking. Johansson [6] has shown in his experiments that observers can recognize walking subjects familiar to them by just watching video sequences of lights affixed to joints of the walker (Fig. 1.1).



**Fig. 1. 1 Marker Positions in Gait Analysis**

However, human gait identification involves challenging issues in measuring and analyzing subject's motion in marker-less approaches due to highly flexible structure and self-occlusion of the human body [2]. Different factors such as footwear, clothing or object carrying, or physical conditions such as pregnancy, leg or foot injuries, or even drunkenness can affect the human's manner of walking. In addition, like most biometrics, gait will instinctively change with age. Besides, variations in the camera viewing angle with respect to the walking subjects can considerably change results. However, a good biometric system should be able to extract the unique characteristics of the biometric, regardless of the presence of such factors [7].

The general framework of automatic gait recognition consists of subject detection, silhouette extraction, feature detection, feature selection and

identification or classification. There are many properties of gait that might serve as recognition features which can be categorized as static features and dynamic features that evolve in time. Static features reflect instantaneous, geometry-based measurements such as stride length. Dynamic measurements, in contrast, are sensitive to the temporal structure of the activity.

In this thesis an automatic feature selection-based gait recognition approach is presented. We propose using GAs on an improved silhouette-based gait recognition approach to search the space of gait pattern feature vectors and demonstrate the importance of feature selection. Feature vectors are represented as eigenvectors in a low-dimensional space, computed by KPCA with the goal of selecting a subset of them which encode important information about the target concept of interest. KPCA is chosen to handle highly complex and nonlinear data distributions, which produces nonlinear subspaces for better feature extraction. GA then is used to select those eigenvectors with higher discriminatory ability to identify the subjects. This is in contrast to the typical strategy of picking a percentage of the top eigenvectors to represent the target concept, independently of the classification task. Our experimental results show considerable Correct Classification (CCR) improvement compared to the results of KPCA feature vectors. This implies that GAs can select eigenvectors encoding mostly information related directly to identity.

As we are going to review most approaches presented for human gait recognition we will observe that the issue of feature selection has been rarely considered in those methods. More specifically, GA has not been used in this context.

We are going to perform GA on our extracted gait patterns and analyze the changes in performance. Since our concentration will be on the effect of feature selection, we have presented a simple model-free approach to extract the required features.

The rest of this thesis is organized as follows: Chapter 2 represents a general overview of two broadly divided categories in gait recognition based on extracted features: model-based and model-free. Chapter 3 reviews briefly the problem of feature selection. Chapter 4 overviews the proposed approach. Experimental results with more details on different sections of the proposed algorithm along with a description on parameters of GA are presented in chapter 5. Finally, chapter 6 concludes the approach and gives some suggestions to improve the current approach in future.

## CHAPTER 2

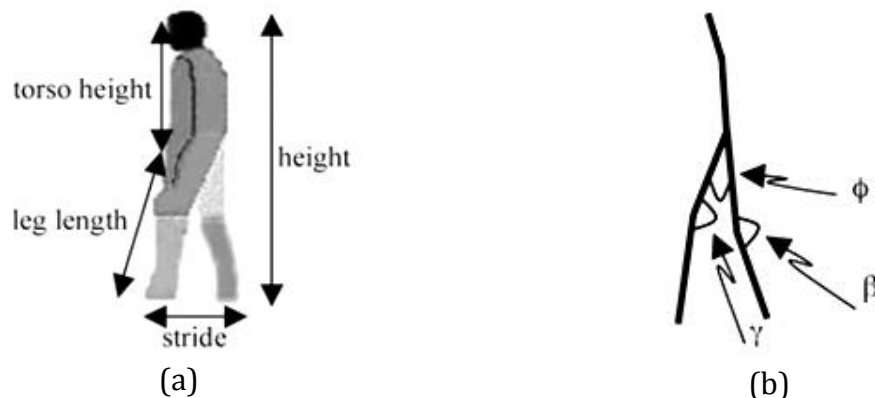
### LITERATURE REVIEW

Approaches presented in order to solve human identification problem using gait as a biometric are generally categorized as model-based and model-free methods based on the type of features they employ.

#### 2.1 Model-Based Approaches

Model-based methods try to explicitly describe subject's movement by modeling or tracking human body components and employ static and dynamic body parameters (Fig. 2.1). Usually, these methods perform model matching in each frame of a walking sequence so that kinematic parameters, such as trajectories, can be measured. Generally, these approaches are scale-independent and view-invariant but sensitive to the quality of gait sequence. Another drawback of approaches in this category is their large computational overhead due to parameters calculations.

A group of these methods consider static structure parameters of body as recognition feature. Bobick and Johnson [8] calculated four distances of human body, namely the distance between the head and foot, head-pelvis, foot-pelvis, and left foot-right foot. They used the four distances to form two groups of static body and stride parameters. Using the within and between condition discrimination power of each set, they could analyze how well a given feature vector will filter identity in a large population.



**Fig. 2. 1 Distinction of (a) Static and (b) Dynamic Parameters**

Yoo [9] constructed a 2D stick figure from human body contour by determining nine coordinates based on human anatomical knowledge. Then, an enhanced back propagation neural network algorithm was employed to identify the gait features extracted from the model and trajectory-based kinematic features calculated from gait sequences. In another work [2], he developed a planar stick figure with six joints and eight sticks extracted from the body contour data. He measured motion parameters to characterize the human gait patterns and features were extracted and selected from the gait sequence by statistical analysis. To classify the gait patterns, a k-Nearest Neighbor (KNN) algorithm was employed. Some other approaches try to employ joint angles trajectories for recognition. Tanawongsuwan and Bobick [10] focused on the trajectories of lower body joint angles from motion capture data. The joint angle trajectories were computed by estimating the offsets between the 3D marker and joints. After normalizing number of footsteps and performing time-normalization on unit-variance signals using Dynamic Time Warping (DTW), recognition was applied using simple nearest neighbor algorithm with Euclidean distance as a measurement criteria. Wang [11] presented a simple and efficient gait

recognition method based on positioning body joints in key frames. The coordinates of joints were computed according to the geometrical characteristics shown while walking. Based on those coordinates the limbs angles were computed on which discrete Fourier transform was applied. At last the nearest neighbor classifier was used to classify subjects having amplitude-frequency and phase-frequency of angles as gait features. Cunado [12] described the periodic motion of thigh as a rotatable line using a Fourier series. By combining temporal Evidence Gathering using a Velocity Hough Transform (VHT) technique with the FS representation of the hip rotation, he could extract a feature-based human gait model from a sequence of images with greater immunity to noise and occlusion. He then used KNN rule for classification having phase-weighted Fourier magnitude information of the motion of the upper leg as feature. A 3D temporal model was proposed by Urtasun and Fua [13] to track and recover motion parameters. Dockstader [14] employed a hierarchical model composed of a set of lines joined at a single point to represent the legs and a periodic, pendulum motion model accompanied by hard and soft kinematic constraints. Using this model, tracking was implemented with a multi-camera Kalman filtering algorithm. Frederic Jean et al. [15] employed an efficient and promising feature which is the spatiotemporal trajectories of the head and feet for view invariant gait modeling. Due to occlusion in feet movement, they estimated the separation between feet in each frame, and determined whether the front foot changes. And then they kept on tracking in each half gait cycle by optical flow. For view normalization, a homograph transformation was computed for each walking

plane to convert it to fronto-parallel view. Some other methods model human body parts separately. In Wang's work [16], a statistical approach based on the Procrustes shape analysis method was employed to obtain a compact representation of the appearance of body shape from the spatiotemporal pattern of the walking action. Besides, human body was modeled as fourteen rigid parts connected to one another at the joints. The tracking results, namely joint-angle trajectories signals, are considered as gait dynamics which were used for identification and verification along with static cues. Boulgouris and Chi [17] extracted body silhouettes and labeled them manually. Based on their experiment each component owned a different discrimination power. The approach they presented combined these components to form a common distance metric for the evaluation of similarity between gait sequences. Using the most contributing parts they could achieve an improvement in recognition performance. Junxia Gu et al. [18] proposed a viewpoint-free marker-less pose recovery and recognition method to automatically extract 3D human joints from the sequence of reconstructed three-dimensional volume data, and then estimated the multiple configuration (normalized joint positions of the whole body and normalized joint positions of the upper limb) and movement features (position, orientation, and height of the body). A Hidden Markov Model (HMM) and an Exemplar-based HMM (EHMM) were then used to model movement and configuration features. A maximum a posteriori (MAP) classifier was then adopted for classification. Zhang et al. [19] presented an approach which involves two generative models, called the kinematic gait

generative model (KGGM) and the visual gait generative model (VGGM), representing the kinematics and appearances of a gait, respectively. These models were synchronized by sharing the same pose manifold during training. Finally, to estimate dynamic gait they employed a new particle filtering algorithm. Trivino et al. [20] analyzed the accelerations produced during a complete gait cycle. Having knowledge about the physiological phases of the human gait, they utilized a fuzzy finite state machine (FFSM) to model the perception of the signal evolution. They divided a gait cycle into four approximately equal phases and the relationship among horizontal acceleration, vertical acceleration, and other indicators were learned. After creating the model, they employed homogeneity, symmetry and the relation weight/legs length as features to represent the individual characteristics.

## **2.2 Model-Free Approaches**

Model-free (or holistic) approaches, on the other hand, usually employ either shape of binary silhouettes or the whole motion of walking person's body, rather than modeling the underlying human body structure or any part of it. Model-free approaches are insensitive to the quality of silhouettes and compared to model-based approaches have a low computational cost. However, their disadvantages are not being robust to viewpoints, scale and changes in illumination and clothing [21, 22].

A baseline algorithm for human identification was proposed in [23] by Sarkar et al. They estimated silhouettes by background subtraction and performed recognition by temporal correlation of silhouettes. Collins [24] established a method

based on matching 2D silhouette templates extracted from key frames for human identification. Using normalized correlation, these frames were compared to training frames. Human identification was performed by nearest neighbor matching among correlation scores. In [25] Lee divided the silhouette into seven regions and presented a representation of human gait appearance based on moments computed in those. He used this moment-based representation for the purpose of person identification and gender classification. Venkat et al. [26] also divided the averaged silhouette into several overlapped parts, including upper, middle, and lower parts as well as left and right parts. To evaluate the impact of these parts on identification, they trained a Bayesian network and achieved promising accuracy with backpack pedestrians. In addition, Li et al. [27] separated the average human silhouette into seven components and then analyzed the impact of each part by performing gait recognition on each part individually, on the entire silhouette and finally on certain combinations of those components. They concluded that the gait of head, arm, trunk and back-leg are important for averaged gait-based human gait recognition. Han and Bhanu [28] employed the idea of motion-energy image (MEI) [29], and proposed the Gait Energy Image (GEI) for individual recognition. This way, instead of considering gait as a sequence of templates, they represented motion in a single image, preserving temporal information. This approach has the advantage of having less storage space and being less sensitive to silhouette noise over using binary silhouette sequence. However, it loses the dynamic variation between successive frames. Individual recognition was then performed by combining statistical gait

features from real and synthetic templates. Liu and Zheng [22] developed the Gait History Image (GHI) originated from GEI to retain temporal information as well as spatial information in form of a pair of templates with time duration of  $1/4$  gait cycle. Chen et al. [30] proposed the frame difference energy image (FDEI) based on GEI and GHI to address the problem of silhouette incompleteness. They first divided a gait cycle into clusters. Then, the average image of each cluster was denoised and the dominant energy image (DEI) was created. By adding the corresponding cluster's DEI and the positive portion of the frame difference between the former frame and the current frame FDEI representation of that frame was obtained. This representation embodied both the kinetic (frame difference) and static (DEI) information of each frame. Xue et al. [31] applied the technique of Infrared (IR) thermal imaging to recognize gaits in order to generate clean silhouettes. They suggested a combination of integral and simplified model to obtain multi-variables gait features. This new model could identify the subject by both analyzing motion information of silhouettes and using anatomic principles. They also exploited wavelet transform (WT), invariant moments and body skeleton theory using GEI to extract gait parameters which were presented to support vector machine for classification. In [16] Wang and Tan proposed a recognition algorithm using spatial-temporal silhouette analysis. They mapped the 2D silhouette sequence into 1D normalized distance signal by contour unwrapping with respect to the centroid of silhouette, to approximate temporal pattern of gait. Then, they used Principal Component Analysis (PCA) on these time-varying signals to do the Eigen-space

Transformation and extract the main components of gait signatures. Finally they employed nearest neighbor classifier (NN) to determine the subject's identity. Hu et al. [32] adopted Gabor filters to create down sampled gray-scale images instead of silhouettes as the original feature. Their approach aimed to consider both spatial feature and temporal relationship. Using Maximization of Mutual Information (MMI) they constructed noise-insensitive gait features. To model the temporal dynamical information of gender related gait patterns and conducting gender classification, Gaussian Mixture Model-Hidden Markov Models (GMM-HMMs) were constructed. Bashir et al. [33] computed optical flow fields for centered and normalized human figures extracted from a complete gait cycle. Gait representation consisted of a Motion Intensity Image (MII) which measured the intensity of relative motion at each pixel location, and weighted sum of four Motion Direction Images (MDIs) each of which represented the likelihood of the direction of motion being along one specific motion direction. These vectors were all discretized to achieve robust features to noise. The Fusion of these vectors resulted in a representation which was discriminative and less sensitive to changes in various covariate conditions including clothing, carrying, shoe, and speed. Some other approaches try to analyze the whole shape of silhouette. BenAbdelkader [34] employed silhouette self-similarity plot (SSP). SSP is defined as the correlation of all pairs of images in the sequence. Normalized feature vectors were extracted from the SSP and used for recognition. Wang et al. [35] represented shape changes of silhouette over time as an associated sequence of vector configurations in a common coordinate system,

which were analyzed using the Procrustes shape analysis approach to extract a structural shape of subject's body shape. For recognition, they adopted supervised pattern classification techniques based on the full Procrustes distance measure. Boulgouris and Chi [36] computed a template using Radon-transformed silhouettes in each cycle of a gait sequence. To construct the feature vector, Linear discriminate analysis (LDA) was then applied on the set of all templates to identify the Radon template coefficients containing most discriminative information. The approach Kellokumpu et al. [37] proposed is based on using a dynamic texture descriptor. They assumed time as the third dimension other than XY axes in the image plane, and thus considered the accumulation of gait sequence as XYT three-dimensional space. 3D local binary patterns (LBP) were used for XYT histogram extraction and represented human movement in a spatiotemporal way. Ran et al. [38] decomposed a video sequence into XT slices and generated periodic patterns called double helical signatures (DHS) by employing iterative local curve embedding algorithm. DHS had the advantage of encoding the appearance and kinematics of human motion and were efficient for recovering gait parameters and detecting simple events. Thus DHS was used for simultaneous segmentation and labeling of body parts in cluttered scenes and load-carrying conditions. In fact, they could infer the carrying status in real time by analyzing the change in symmetries of periodic patterns. In [39], Chen et al. employed the Factorial HMM (FHMM) as feature-level fusion scheme to fuse different gait features which was compared with the Parallel HMM decision-level fusion scheme. FHMM has a multiple-layer structure and combines the information

from each layer to obtain the model parameters. PHMM on the other hand, has no interlayer connections and each layer runs independently as an HMM-based classifier.

Regardless of developing a model-based or model-free approach, the extracted features can be high dimensional. Having a large number of features might represent the target concept better, but without employing a feature selection strategy, however, many of them could be either redundant or irrelevant to the classification task [40]. Features extracted from segmented video sequences are commonly not effective for classification and due to their high dimensionality, too many training samples will be essential. Such high-dimensional feature spaces present several problems to gait recognition. First, it may affect the performance of many classification methods. Second, due to the existence of many gait-related factors, e.g., health, age, body size, weight, speed, etc., coupled with limited understanding of the underlying recognition mechanism, many redundant features tend to be included to avoid any loss of useful information [41]. Hence, a feature selection method in such a situation would be useful to provide valuable clues to extract the underlying distinctness among human gait patterns. Generally speaking, the choice of features to represent the patterns affects several aspects of the recognition problem such as accuracy, required learning time and necessary number of samples. Although all the methods mentioned above have presented notable recognition rates using their own feature vector, but beyond a certain point, the inclusion of additional features would lead to a decrease in performance. This

apparent paradox presents the feature subset selection problem in automated design of pattern classifiers. In other words, it is desired to identify and select a useful subset of features to represent patterns from a larger set of often mutually redundant or irrelevant features. The main goal of feature subset selection, thus, is to reduce the number of features used in classification while maintaining acceptable classification accuracy.

Feature subset selection in the context of practical applications such as gait recognition presents a multi-criterion optimization function, for which we are going to consider both the number of features and accuracy of classification. Automatic feature subset selection and in particular, GAs offer a particularly attractive approach for such problems since they are generally quite effective for rapid global search of large, non-linear and poorly understood spaces. Moreover, GAs are very effective in solving large-scale problems [42]. Examples include face detection [40, 43], gender classification [9, 44], vehicle detection [40, 43], seed discrimination [45], target detection [46], tracking [47], and video categorization [48].

The use of feature selection, however, has not been given enough consideration in gait recognition. These approaches [2, 28 and 49] have mainly considered conventional dimensionality reduction or statistical tools, such as PCA and ANalysis Of VAriance (ANOVA). In [41] Guo and Nixon used a sequential selection method based on mutual information (MI). By evaluating the statistical dependence between two random variables and creating a relation with the Bayes classification error, MI can be used to measure the utility of selected features in classification of human gait

data. Begg et al. [50] employed a Hill-climbing feature selection algorithm to choose a subset of features on which Support vector machine (SVM) was applied for automated recognition of gait pattern changes due to ageing. In another work [51] they could improve their performance by using a forward feature selection algorithm in which a feature was sequentially added one at a time that most increased or least decreased the classification accuracy. Bashir et al. [52] developed both supervised and unsupervised feature selection methods in their approach in order to extract the most informative shape and dynamic features from GEI as the gait representation. For supervised gait feature selection, a cross validation based approach and a priori knowledge about the GEI feature characteristics were adopted. For unsupervised feature selection, they proposed an effective measurement of the relevance of GEI features without assuming that a labeled dataset was available.

## CHAPTER 3

### FEATURE SELECTION

The selection of an optimal subset of features is a necessary and important step in pattern recognition; often a large number of features are extracted to better represent the target concept. Given a set of features, the problem is to identify a powerfully predictive subset of features within the database that leads to the smallest classification error. This is essentially an optimization problem that involves searching the space of possible feature subsets to find one that is optimal or near-optimal with respect to a certain criterion [40]. Without employing some kind of feature selection strategy, however, many of them could be either redundant or even irrelevant to the classification task. As a result, the classifier might not be able to generalize nicely. In fact, by extracting as much information as possible from a given data set while using the smallest number of features, significant computation time is saved and models which generalize better for unseen data points are built. Besides, the choice of features used to represent patterns affects the number of examples needed for learning. Feature selection is a difficult task because it requires searching through a space which may be high-dimensional. An exhaustive search is computationally prohibitive especially when there are a large number of features. This has led to the development of a wide range of feature selection methods.

There are many algorithms that use a greedy search through the solution space. Decision tree algorithms such as Quinlan's ID3 [53], C4.5 [54] and CART, are some of

the most successful supervised learning algorithms. Michalski proposed the AQ learning algorithm. Narendra and Fukunaga presented a Branch and Bound algorithm [55]. A well-known algorithm that relies on relevance evaluation is RELIEF. Subset search algorithms [56] search and capture the goodness of each subset. There are again many algorithms that employ exhaustive, heuristic and random search. Clustering algorithms are also used for feature selection process for which ROCK and CACTUS are a few of them. Among these, Naive Bayes or Bayes' Rule is the basis for many machine-learning and data mining methods [55].

Generally, feature subset selection algorithms can be classified into two categories based on whether or not feature selection is performed independently of the learning algorithm used to construct the verifier. If feature selection is done independently of the learning algorithm, the technique is said to follow a filter approach. Otherwise, it is said to follow a wrapper approach.

The first one is computationally more efficient but its major drawback is that an optimal selection of features may not be independent of the representational biases of the learning algorithm that is used to build the classifier. On the other hand, the wrapper approach involves the computational overhead of evaluating a candidate feature subset by executing a selected learning algorithm on the database using each feature subset under consideration [42].

For instance, sequential direct selection starts by selecting the best feature among the whole set of available ones. At each step, it keeps the latest selected subset of features and adds to it the best feature among the remaining ones. In

contrast, sequential backward selection starts by selecting the complete set of features and rejects one feature at a time. At a given step, the feature which causes the smallest decrease of the performances is permanently removed. The process stops when the removal of each of the remaining features causes a high decrease of performances [45]. The main weakness of these methods is that they may reject a variable at an early step of the search process which is then unavailable for a later possible improvement in performance. This means that this kind of feature selection methods may lock at a local minimum during the search process.

An alternative is the use of GAs as an effective random search approach of wrapper model, in which features that are removed in successive steps of selection, may simply also be removed from the representing chromosomes. That way the optimal solutions found for one feature subset are reused in the next step.

### **3.1 GA-Based Feature Selection**

GAs which are able to explore, in an efficient way, high dimensional spaces about which little is known a priori, for function optimization [57]; encoding important information about the target concept of interest. GAs, a form of inductive learning strategy, are adaptive search techniques based on a direct analogy to Darwinian natural selection and genetics in biological systems, which have demonstrated substantial improvement over a variety of random and local search methods.

GAs work with a set of candidate solutions called a population. Using the Darwinian principal of “survival of fittest”, GA generates successive population of

alternate solutions containing chromosomes, until after a series of generations it obtains optimal solution. Associated with the characteristics of exploitation and exploration search, GA can handle large search spaces efficiently, and thus there is less probability of getting local optimal solution. A fitness function evaluates the quality of each solution based on which chromosomes with higher probability will be selected for reproduction. The crossover and mutation functions are the main operators that randomly impact the fitness value. Offspring replaces the old population based on a specific strategy and the population for next generation is formed [58].

Since GAs is basically a domain independent search technique, they are ideal for applications where domain knowledge and theory is difficult or impossible to provide. The major reason for GAs popularity in various search and optimization problems is its global perspective, wide spread applicability and inherent parallelism. GA is based on the observation that the evolution of natural species is very efficient at adapting to changing environments. By simulating the evolution process, GAs may provide a good way for optimizing artificial systems [59].

Siedlecki et al. [60] presented one of the earliest studies of GA-based feature selection in the context of K-nearest-neighbor classifiers. They used GA to find an optimal binary vector, where each bit was associated with a feature. GAs have been used to search for feature subsets in conjunction with several classification methods. Reymer et al. [61] implemented a GA-KNN classifier for identification of favorable water-binding sites on protein surfaces. Yang et al. [62] proposed a

feature selection approach using GAs and NNs for classification. A standard GA with rank-based selection strategy was used. They tested their methods using several benchmark real-world pattern classification problems and reported improved results. However, they used the accuracy on the test set in the fitness function, which is not appropriate since it introduces bias to the final classification. Tsai [63] developed an ad hoc iterative variable reduction algorithm for a probabilistic neural network (PNN) to identify noise and redundant variables. This iterative approach employed a weighted PNN with one smoothing factor for each variable in the variable reduction phase. A basic PNN was developed finally based on the subset of selected variables. In [64] Jiang et al. employed GA-based feature selection whose fitness function combined the number of features to be used and the error rate of Bayesian classifier. Huang et al. [65] presented a feature optimization approach based on GA and SVM classification accuracy. Chtioui et al. [66] investigated a GA approach for feature selection in a seed discrimination problem. Using standard GA operators, they selected the best feature subset from a set of 73 features. Vafaie et al. [67] conducted a comparison between important score (IS), a greedy-like feature selection method and GAs. They represented the feature selection problem using binary encoding and standard GA operators. The evaluation function was solely based on classification performance. Using several real world problems, they found that GAs are more robust at the expense of more computational effort.

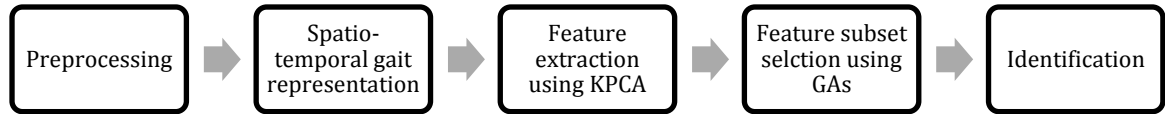
## CHAPTER 4

### METHOD OVERVIEW

The goal of the approach we have presented here is to choose the best feature vector with highest discrimination among a large number of features for every person extracted from his silhouette contour projections. This is accomplished through exploiting GA on eigenvectors of four contour projections of the silhouette of person accumulated over the whole sequence. Fig. 4.1 shows the overall framework of the approach.

The first stage, preprocessing, includes noise reduction, background subtraction and formation of silhouette's global temporal accumulation from the input video sequences. These steps result in static parameters such as approximate height of subject, his velocity and period of walking, based on which the pattern will be normalized in other steps. The spatio-temporal information is represented in a single 2D gait template by using multi-projection of preprocessed silhouettes. Feature vectors are then constructed using eigenvectors derived from applying KPCA on the gait patterns to project the data in a lower-dimensional space. However, these feature vectors will not be employed in classification. In fact, in the following step feature selection using GA is performed with goal of choosing a subset of eigenvectors, encoding mostly important information about the target concept of interest. Finally, the evaluation function employed here contains two terms, the first part is based on classification accuracy of a validation set and the

second portion belongs to the number of eigenvectors selected. Finally, KNN or Naive Bayes Classifiers are used to identify subjects in the test set.



**Fig. 4.1 Proposed Algorithm Framework**

Based on physiological research results, most of the walking dynamics take place in the sagittal plane, or the plane which vertically bisects the human body [68]. Thus, our method assumes a single subject moving at a constant speed, fronto-parallel to the camera, against a static background. However, this approach can be generalized to more practical situations.

## **4.1 Gait Pattern Extraction and Representation**

### **4.1.1 Preprocessing**

In this phase a kernel regression method [69] is first applied on frames of each sequence to remove noise and retain the most representative information of each video. Then, a background subtraction approach based on using the minimum graph cut method [70] is employed to segment the foreground, resulting in qualitatively and quantitatively cleaner segmentations. The silhouettes extracted in this process will be later used in gait pattern representation. In the next step, Sobel operator is applied on all frame images; a combination of this data and the extracted silhouettes in previous step results in removing all static objects, leaving only edges belonging to moving objects. A velocity filtering algorithm [70] is then applied on these edges to determine the velocity and approximate height of the subject. The last step in this

phase would be period estimation based on silhouette width variation. These steps are shown in Fig. 4.2 and will be discussed in details as follows.

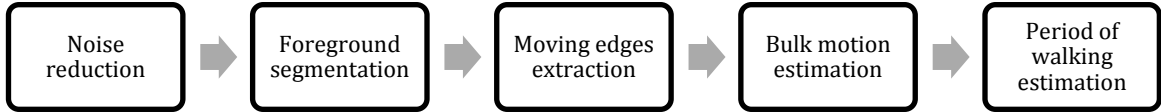


Fig. 4.2 Preprocessing Framework

#### 4.1.1.1 Noise Reduction

The input frames are pre-processed using a kernel regression method [69] to recover the noiseless high-frequency information corrupted by limitations of the imaging system. In order to achieve this goal the approach includes two steps, where first an initial estimate of the image gradient is made using a second order kernel regression method. Next, this estimate is used to measure the dominant orientation of local gradients in the image. Then, in a second filtering stage, this orientation information is used to adaptively “steer” the local kernel, resulting in elliptical contours spread along the directions of the local edge structure. With these locally adapted kernels, the denoising is affected most strongly along the edges, rather than across them, resulting in strong preservation of details in the final output. The data adapted kernel is of the form  $K_{H_i^{steer}}(x_i - x)$  where  $H_i$ s are steering matrices defined as  $H_i^{steer} = h\mu_i C_i^{-1/2}$  where  $C_i$ s are symmetric covariance matrices based on difference in the local grey-values [69].

Since steering kernel regression is most effective when an iterative regression/denoising procedure is used to exploit the less noisy image of each iteration to estimate the radiometric terms of the kernel in the next iteration,

estimated kernels are directly applied on the non-interpolated samples which results in the denoised image in first iteration. Further iterations apply the modified kernels on the denoised pixels leading to more aggressive noise removal.

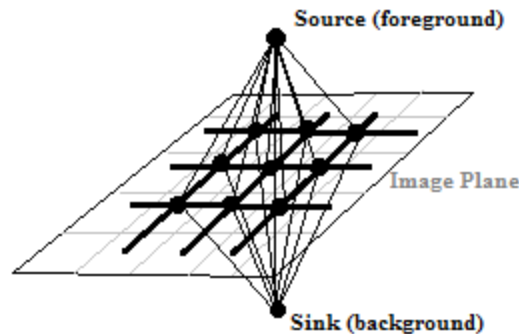
#### **4.1.1.2 Foreground Segmentation**

The preprocessing phase is followed by background subtraction. The method used here for foreground segmentation is based on using the minimum graph cut method proposed in [71]. In this approach a model is constructed after each frame in the video stream, either offline or dynamically updated based on current frame. To build the background model the pixel color from every fourth frame is taken and the data above and below a pair of thresholds is thrown away. From the remaining numbers it estimates the mean and variance of each pixel's color, assuming a normal distribution.

Then next frame is compared to the model in a pixel-based basis, and extract those sufficiently different as part of foreground. To achieve this goal, a graph is built based upon the image in which each vertex links to exactly six other nodes: the foreground and the background, plus the vertices of its four-connected neighbors (Fig. 4.3). The weights of the links are derived directly from the differences measured between the current frame and the background model at the corresponding pixel.

Finally using standard methods based upon graph flow an optimal cut, with minimum cost, would be found which separates the foreground from the

background. The nodes still connected to the foreground node will be labeled as foreground and the rest are specified as background.



**Fig. 4.3** Constructed Graph for a Sample 3×3 Portion of the Image Plane [71]

It is also possible to create the graph representing all frames of video at once. In such case each pixel vertex would have four previous connections plus two temporal connections. The reconstruction, however, remain the same.

#### **4.1.1.3 Static Parameters Estimation**

In the next step, a velocity filtering algorithm is employed to determine the bulk motion of the silhouette of person. As mentioned earlier this algorithm will be applied on moving edges of each frame. These edges are extracted as the common output of background subtraction and Sobel edge detection. Using this motion information, a global temporal accumulation describing the person's average shape is formed over the gait sequence. Here we use a velocity filtering algorithm [70] to determine the bulk motion of the person over gait sequence independent of shape. This accumulation will be later employed to extract size and shape of the person's body segments. To perform the global temporal accumulation the following formula is used.

$$A_v(i, j) = \sum_{n=0}^N I_n[i + v * n, j] \quad (1)$$

Where  $A_v$  is the accumulation for velocity  $v$  (in pixels per frame),  $I_n$  is the image intensity function at frame  $n$ ,  $i$  and  $j$  are coordinate indices and  $N$  is the number of frames in the gait sequence.

The algorithm sorts the objects in the scene according to their velocity and starting position, producing an accumulation for each possible object velocity. In fact, each object in scene accentuates the accumulation it belongs to base on its edge strength and number of frames it has been observable in the whole sequence (Fig. 4.4). Another important factor is that how close the velocity of object has been to the velocity of accumulation. In other words, the highest peak in the plot of maximal intensity versus velocity indicates the object's velocity. If there is more than one moving object in the scene with a similar velocity, some knowledge of person's shape is required to be able to distinguish its velocity.



**Fig. 4. 4 Global Temporal Accumulation [70]**

Having an approximate height and velocity for each subject, the last static parameter to be estimated is the period of walking. The motion of a person's limb during normal gait creates a complex periodic pattern, composed of many different

components which can be approximately modeled by a single sinusoid. The gait frequency and phase are particularly useful components since they describe this motion to a large extent and can be easily extracted without resolving limb dynamics. We have chosen simple features of the silhouette, such as variation of its width, as a measure of period calculation to decrease computational overhead.

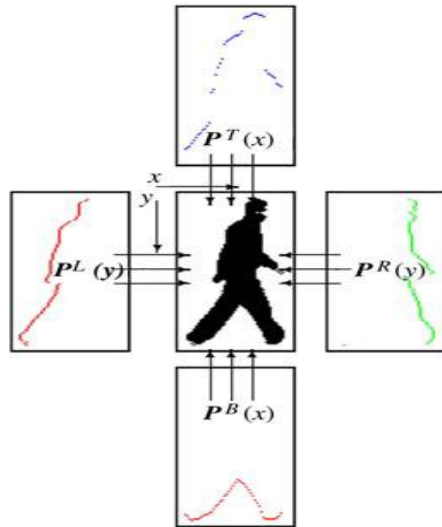
#### **4.1.2 Gait Representation**

To create gait patterns we have used an improved spatio-temporal gait representation based on the one proposed in [72]. This pattern is obtained by using multi-projections of silhouettes.

Input silhouettes are first normalized based on the person's approximate height gained through global temporal accumulation of previous phase. In fact each silhouette is resized proportional to a uniform bounding box and then they are horizontally aligned so that the upper half silhouette part is centered with respect to its horizontal centroid.

Then, projections of silhouettes, from a sequence of binary silhouette images are extracted from which gait patterns will be created. These binary silhouettes,  $B_t(x,y)$ , are indexed spatially by pixel location  $(x,y)$  and temporally by time  $t$ . The distance vectors are the differences between the bounding box and the outer contour of silhouette. This definition includes four different images for top, bottom, left and right distance vectors. A sample silhouette along with its distance vectors are displayed in Fig. 4.5. The distance vectors are the differences between the bounding box and the outer contour of silhouette. In other words, the values in top

and bottom contour projection vectors are the number of rows between bounding box and silhouette at each column which makes their width the same as the width of bounding box. The same thing holds for left and right distance vectors, and gives them the same height as bounding box. This process thus creates four new 2D images referred to as gait patterns.



**Fig. 4.5 Silhouette Representation [72]**

The gait pattern image for top-projection is formulated as  $P^T(x, t) = \sum_y B_t(x, y)$  where each column (indexed by time  $t$ ) is the top-projection of silhouette image  $B_t(x, y)$  which means the count of the number of rows between the top side of the bounding box and the outer contour in the same column. This way a 2D spatio-temporal pattern is formed by stacking row projections together. Other projections  $P^B(x, t)$ ,  $P^L(y, t)$  and  $P^R(y, t)$  are constructed in the same way by stacking rows/columns from bottom, left and right projections respectively. The variation of each component of the distance vectors can be regarded as gait signature of that object [72].

## 4.2 Feature Extraction using Kernel PCA

In this step, we are going to perform a dimensionality reduction procedure on the gait patterns extracted in previous step. As a result, training gait features that form feature database are obtained. This is independently repeated for each gait pattern produced from the projections.

Conventional linear subspace methods such as principal component analysis (PCA) can only produce linear subspace feature extractors. These are unsuitable for highly complex and nonlinear data distributions. In contrast, kernel subspace methods such as KPCA can capture higher order statistics present in a dataset, thus producing nonlinear subspaces for better feature extraction.

In principle, kernel methods map the input data to a higher dimensional feature space where conventional linear subspace methods can be used, with the resulting subspaces being nonlinear with regards to the original input space. In essence, we can perform PCA in feature space and create result of nonlinear PCA in input space. Experiments and comparisons have shown that KPCA almost always outperforms PCA [73].

To apply KPCA on a data matrix  $a = [x_1 \dots x_n] \in R^{m \times n}$ , we map it to a higher dimensional space  $F$  using a kernel function  $\varphi: R^m \rightarrow F$  and then perform a linear PCA in  $F$  [73].

Using  $\varphi$ ,  $a$  is transformed into  $A = [\varphi(x_1) \dots \varphi(x_n)]$ . Thus the covariance matrix in this new space  $F$  is

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n \varphi(x_i) \varphi(x_i)^T \quad (2)$$

Besides, the map  $\varphi$  is induced by a kernel function  $k(\cdot, \cdot)$  that allows efficient evaluation of inner products of vectors in higher dimensional space  $F$ :  $k(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$ . The most often used kernel functions are Gaussian kernel, polynomial kernels and sigmoid kernels [72]. We employed Gaussian kernel which is defined as

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (3)$$

Considering the matrix  $M = \hat{A}^T \hat{A}$  and its eigenvalue decomposition,  $M = Q \Delta Q^T$  by using  $k(\cdot, \cdot)$ ,  $A^T A$  can be evaluated without having to perform the mapping  $\varphi$  since  $A^T A$  contains only dot products between the  $\varphi(x_i)$ s. Matrix  $M$  is the kernel matrix for KPCA. Via SVD, the rank- $r$  singular value factorization of  $\hat{A}$  is

$$\hat{A}^r = \left[ \hat{A} Q^r (\Delta^r)^{-\frac{1}{2}} \right] \left[ (\Delta^r)^{\frac{1}{2}} \right] \left[ (Q^r)^T \right] \equiv U^r \Sigma^r (V^r)^T \quad (4)$$

$U^r$  is defined implicitly by linear expansion of mapped input data

$$U^r = A \hat{v} Q^r (\Delta^r)^{-1/2} = A \alpha \quad (5)$$

Where  $\alpha = \hat{v} Q^r (\Delta^r)^{-1/2}$  contains the expansion coefficients. The reconstruction of  $A$  using the first- $r$  kernel principal components would be  $A^r = \hat{A}^r + \mu_A$  [67].

Usually we need to keep a smaller number of eigenvectors corresponding to the largest eigenvalues. It, however, has been found in several studies that different eigenvectors encode different kind of information [40]. In essence, different tasks make different demands in terms of the information that needs to be processed, and this information is not contained in the same ranges of eigenvectors.

The number of features which can be safely introduced in a predictive model is dependent on the number of learning patterns. Only a few features are necessary to correctly describe a data set including a small number of learning patterns. With a limited number of learning patterns, adding new features to a pattern recognition system may degrade rather than increase the performances. Hence, feature selection is necessary to separate adequately qualitative classes with a limited number of learning patterns.

Thus, we apply GA to search the space of eigenvectors with the goal of selecting a subset of them which encode important information about the manner of walking for each individual.

### **4.3 Genetic Feature Subset Selection**

Evolutionary algorithms offer a particularly attractive approach to multi-criteria optimization because they are effective in high-dimensional search spaces. If the number of variables is not too large, an exhaustive search could be suitable, but in optimization problems with a considerable amount of features, this is unfeasible.

A GA is a model of machine learning which derives its behavior from a metaphor of some of the mechanisms of evolution in nature. This is done by the creation within a machine of a population of individuals represented by chromosomes. The individuals represent candidate solutions to the optimization problem being solved. In GAs, the individuals are typically represented by n-bit binary vectors. The resulting search space corresponds to an n-dimensional Boolean space. It is assumed that the quality of each candidate solution can be evaluated using a fitness

function. In other words, in this kind of optimization problems, a fitness function ( $x$ ) should be maximized or minimized over a given space  $X$  of arbitrary dimension. The main issues in applying GAs to any problem are selecting an appropriate representation and an adequate evaluation function.

GAs use some form of fitness-dependent probabilistic selection of individuals from the current population to produce individuals for the next generation. The selected individuals are submitted to the action of genetic operators to obtain new individuals that constitute the next generation. The genetic operators transform the population through successive generations, extending the search until arriving to a satisfactory or optimum result. It is very important that the population be diverse and keeps characteristics of adaptation acquired by the previous generations [74]. Mutation and crossover are two of the most commonly used operators that are used with GAs that represent individuals as binary strings. Mutation operates on a single string and generally changes a bit at random while crossover operates on two parent strings to produce two offspring. Other genetic representations require the use of appropriate genetic operators. The process of fitness-dependent selection and application of genetic operators to generate successive generations of individuals is repeated many times until a termination criterion is satisfied. This termination criterion can be defined as reaching a predefined time limit or number of generations or population convergence or satisfactory solution is found.

In practice, the performance of GA depends on a number of factors including: the choice of genetic representation and operators, the fitness function, the details

of the fitness-dependent selection procedure, and the various user-determined parameters such as population size, probability of application of different genetic operators, etc.

#### **4.3.1 Feature Selection Encoding**

Each individual in the population represents a candidate solution to the feature subset selection problem. In the feature selection problem the main interest is in representing the space of all possible subsets of the given feature set. Then, the simplest form of representation is binary representation where each feature in the candidate feature set is considered as a binary gene and each individual consists of fixed-length binary string representing some subset of the given feature set. An individual of length  $l$  corresponds to an  $l$ -dimensional binary feature vector  $X$ , where each bit represents the elimination or inclusion of the associated feature. Then,  $X_i = 0$  represents elimination and  $X_i = 1$  indicates inclusion of the  $i$ th feature [45].

#### **4.3.2 Fitness Evaluation**

The fitness evaluation is a mechanism used to determine the confidence level of the optimized solutions to the problem. Usually, there is a fitness value associated with each chromosome, e.g., in a minimization problem, a lower fitness value means that the chromosome or solution is more optimized to the problem while a higher value of fitness indicates a less optimized chromosome.

Choosing an appropriate evaluation function is an essential step for successful application of GAs to any problem domain. In order to use GAs as the search procedure, it is necessary to define a fitness function which properly assesses the

decision rules generated. Each subject is classified using the current feature subset. If this is the appropriate classification, then the subject's identity has been recognized correctly. Classification accuracy, the number of selected features, and the feature cost are the three criteria generally used to design a fitness function. Thus, for the chromosome with high classification accuracy, a small number of features, and low total feature cost a high fitness value should be produced. Here the overall fitness function will be evaluated by adding the weighted sum of the match score of a bit string indicating present features along with the number of features exploited meanwhile; feature cost has been considered to be equal for all the features. We solve the multiple criteria problem by creating a single objective fitness function that combines all of the goals into one. As defined by formula 6, the fitness has two predefined weights: (i)  $W_A$  for the classification accuracy; (ii)  $W_F$  for the summation of the selected feature (with nonzero value in representation string). However, the former part is the major concern, and that is why  $W_A$  gets a larger value. The final fitness function is defined as:

$$\text{Fitness} = W_A \times \text{Correct Recognition Rate (CRR)} - W_F \times \frac{\text{Number of selected features}}{\text{Chromosome length}} \quad (6)$$

The match score is evaluated based on feedback of some classifiers which try to classify the input database considering the current available features. The chromosome with high fitness value has high probability to be preserved to the next generation, so the above settings should be appropriately defined according to the requirements of system.

### **4.3.3 Initial Population**

Generating an initial population of chromosomes is often achieved at random, but the population may be initialized by chromosomes which are already known to perform well. Here to generate the initial population we have created a random number for each chromosome to define the number of present features. These value, also are scattered randomly through the whole chromosome. So we would have a population with different permutations and number of features. When random initialization of binary chromosomes is used, each bit of the chromosomes is randomly set to 0 or 1 according to a probability which is called the initialization probability.

### **4.3.4 Selection**

The intention of the selection operator is to select parents that increase the probability to reproduce members of the population that have good values of the fitness function. There are some methods for the genetic selection. The choice of an adequate mechanism of selection can save significant computation time [74]. We have employed two selection techniques to compare the results, elitism and linear ranking.

Elitism reserves one (or a few) slots in the next generation for the highest scoring chromosome of the current generation, without allowing that chromosome to be crossed over in the next generation. In one of those slots, the elite chromosome will also not be subject to mutation in the next generation. This method can very rapidly increase performance of GA, because it prevents losing the

best found solution.

In ranking selection each individual in the population is assigned a numerical rank based on fitness (probabilities of ranked individuals are linearly weighted in linear ranking case), and selection is based on these rankings rather than absolute differences in fitness. The advantage of this method is that it can prevent very fit individuals from gaining dominance early at the expense of less fit ones, which would reduce the population's genetic diversity and might hinder attempts to find an acceptable solution.

#### **4.3.5 Operators**

To explore the whole space of features crossover and mutation operators must be employed. Therefore, two individuals from the whole population of individuals are selected. The selection is dependent on the value of the fitness function of each individual. The well-adapted individuals have a greater chance at being selected.

Crossover is the operator responsible for the recombination of characteristics of the parents during the reproduction, enabling the next generations to inherit these characteristics. It is considered the most important genetic operator [74]. The crossover operator exchanges genes between two chromosomes using different techniques. In the proposed method a uniform crossover is exploited to avoid destroying the schema in case of presence of dependency among neighbor eigenvectors.

The mutation operators, on the other hand, are necessary to guarantee a genetic diversity of the population; they do this by randomly modifying one or more

components of a chosen structure. For instance, in binary coded genes, this is done by changing genes codes from 0 to 1 or vice versa [74]. Mutation makes possible the introduction of new elements in the population. The mutation operator is applied to the individuals with a probability given for mutation rate; in general a small mutation rate is used in practice. Here, we have employed the bitwise mutation.

#### **4.4 Gait Recognition**

Feature extraction can be used in conjunction with numerous methods for pattern classification. The well-known statistical methods can be divided into two general classes. A classifier from each category has been utilized to evaluate chromosomes during fitness estimation in this approach which are K-Nearest Neighbor and Naive Bayes Classifiers.

Among the various methods of supervised statistical pattern recognition, the Nearest Neighbor rule achieves consistently high performance, without a priori assumptions about the distributions from which the training examples are drawn. It is carried out under the assumption that the similar images belong to the same category. It involves a training set of both positive and negative cases. A new sample is classified by calculating the Euclidean distance to the nearest training case; the sign of that point then determines the classification of the sample [75]. The k-NN classifier extends this idea by taking the k nearest points and assigning the sign of the majority. Larger k values help reduce the effects of noisy points within the training data set, and the choice of k is often performed through cross-validation.

Given a set of d instance-label pairs  $(X_i, L_i)$ ,  $i = 1, 2, \dots, d$ , where  $X_i \in \mathbb{R}^n$ ,  $L_i$  is the

category label of  $X_i$ . Each feature vector represents a point in an  $n$ -dimension feature space and is used as a query to compute the “closeness” to the other feature vectors.  $K$ -NNs that are the closest to the query are returned. The query image is assigned with the most common category among its  $k$  nearest neighbors. “Closeness” is defined in terms of similarity measure [75]. There are several similarity measures based on common distance functions such as Euclidean, Mahalanobis and etc. We use Euclidean distance, where the Euclidean distance between two points  $X$  and  $Y$  is defined as the following:

$$d(X, Y) = \sqrt{\sum_{k=1}^m \sum_{l=1}^n (x_{k,l} - y_{k,l})^2} \quad (7)$$

The Naive Bayes Classifier technique is based on the so-called Bayesian theorem and is particularly suited when the dimensionality of the inputs is high. Despite its simplicity, Naive Bayes can often outperform more sophisticated classification methods. This method assumes that the form of the class-conditional density function of the features is known in advance. For example, it is commonly assumed that the features follow a multivariate Gaussian distribution. The training data are used to estimate the parameters of these class-conditional densities (e.g., the mean vector and the covariance matrix for the Gaussian distribution), and these estimated densities are then used to classify test patterns. In simple terms, a Naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. Even if features depend on each other or upon the existence of other features, a Naive Bayes classifier

considers all of these properties to independently contribute to the probability.

Given a set of variables,  $X = \{x_1, x_2, \dots, x_d\}$ , we want to construct the posterior probability for the event  $C_j$  among a set of possible outcomes  $C = \{c_1, c_2, \dots, c_d\}$ . Using Bayes' rule:

$$p(C_j | x_1, x_2, \dots, x_d) \propto p(x_1, x_2, \dots, x_d | C_j) p(C_j) \quad (8)$$

Where  $p(C_j | x_1, x_2, \dots, x_d)$  is the posterior probability of class membership, which means the probability that  $X$  belongs to  $C_j$ . Since Naive Bayes assumes that the conditional probabilities of the independent variables are statistically independent we can decompose the likelihood to a product of terms:

$$p(X | C_j) \propto \prod_{k=1}^d p(x_k | C_j) \quad (9)$$

and rewrite the posterior as  $p(C_j | X) \propto p(C_j) \prod_{k=1}^d p(x_k | C_j)$ . Using this Bayes' rule, we label a new case  $X$  with a class level  $C_j$  that achieves the highest posterior probability.

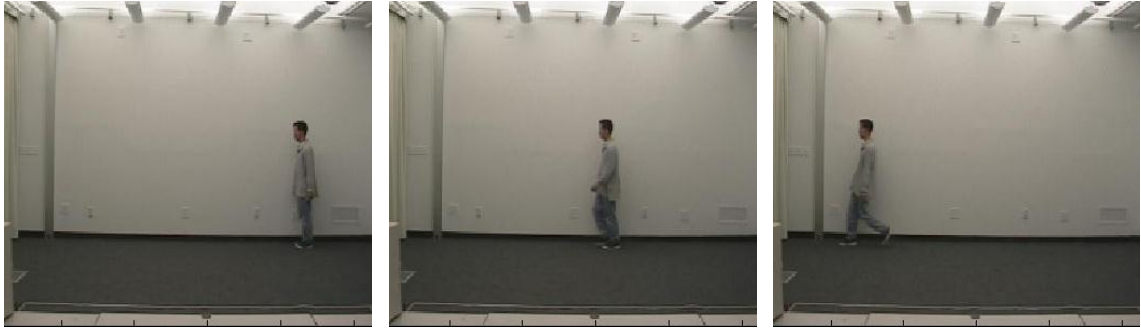
## CHAPTER 5

### EXPERIMENTAL RESULTS

In this chapter we are going to analyze the performance of the approach proposed in previous chapter on a standard database. In order to do this, we first review the outcome of each step of method separately and then evaluate the discrimination ability of our approach compared with some other major gait recognition methods.

#### 5.1 Dataset

The proposed approach has been tested on the Georgia Tech (GT) database [8], which consists of 268 sequences from 20 subjects (6 female/14 male). These sequences are collected under two viewing conditions: side view and angle view (45°). For each subject and angle condition, there are 6 trials. Data collection is done indoors and outdoors. Considering these conditions, we use the indoors data, where lighting is at a constant level and subjects move with approximately constant velocity in front of a plain static background. The sequences were taken at different time intervals and the subject's distance to the camera is changing. This data subset consists of 108 sequences (6 sequences for 18 subjects). Each video sequence is stored in digital video (DV) format, encoded in color PAL format at a resolution of 320×240 pixels, recorded at a rate of 29.97 frames per second. Each sequence typically consists of 80–120 frames, or around 3 full gait cycles. First, middle and last frames of two sample sequences are displayed in Fig 5.1.



(a) Selected Frames of Sequence side0001m03



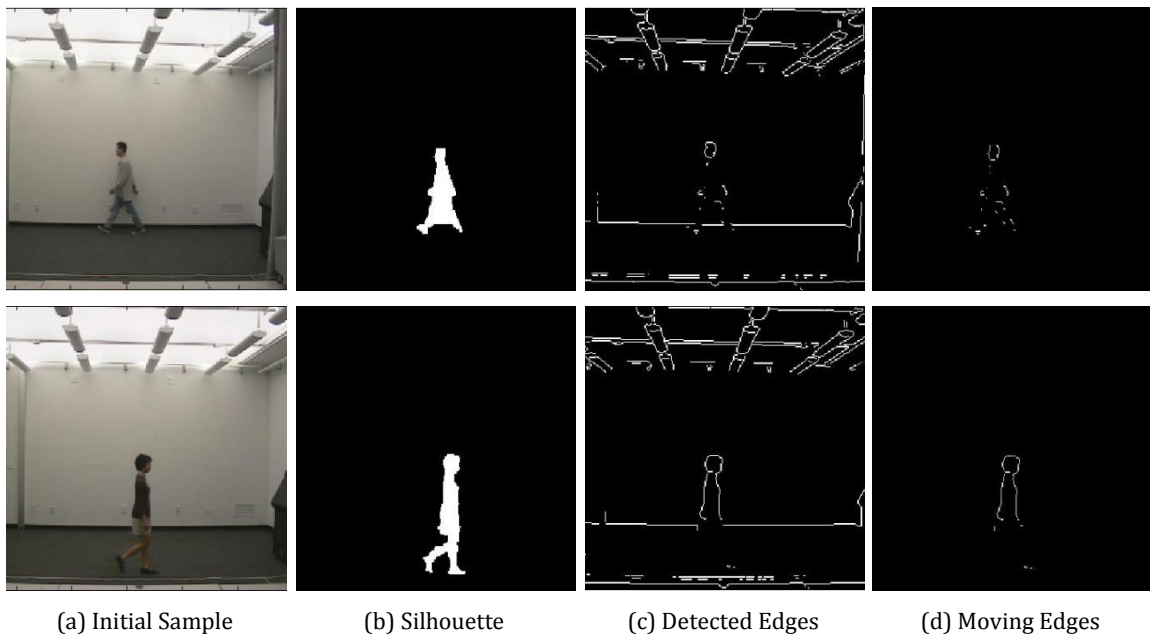
(b) Selected Frames of Sequence side0019f02

**Fig. 5. 1 Sample Sequences of GT Database.**

## 5.2 Algorithm in Steps

As discussed earlier, before performing the foreground extraction phase, we should build a model of background based on which the graph and its vertices are created for each pixel and corresponding weights are calculated. Here we have employed a static model instead of dynamic model which gets updated after each frame. In simple activities like walking static models perform with the same performance of dynamic ones; besides, they have less computational overhead on the system. Segmenting the graph using a standard graph-cut algorithm appears better than approaches using morphological operations, at overcoming the effects of noise by aggregating information from a local neighborhood around each pixel, while remaining true to the underlying data.

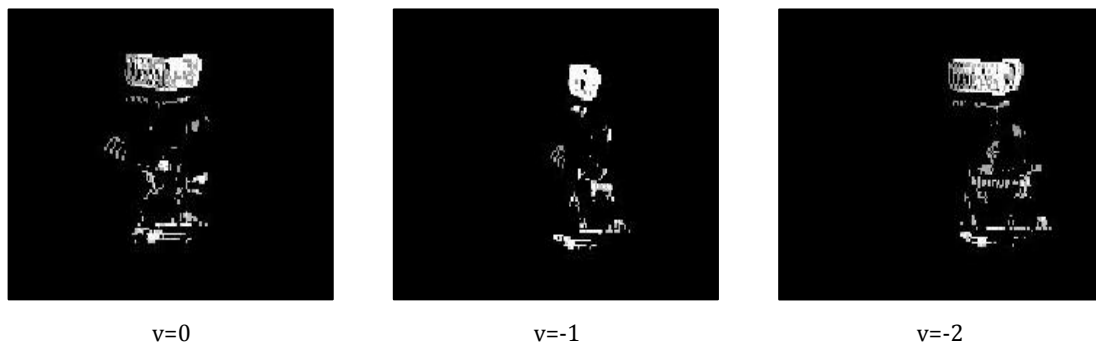
After subtracting the background, edges of each frame are extracted using the Sobel operator. Using the information of moving silhouette, we can now extract the moving edges. This stage is a preparation for estimating the subject's walking speed which was elaborated in previous chapter. Typical results of these two steps are shown in Fig. 5.2 for two input instances. We can see in the pictures that in some parts of the body the algorithm has not been able to detect the foreground correctly. However, this problem happens only in frames where the intensity of pixels belonging to foreground is too close to the ones constructing background. We will explain later that considering the periodic nature of gait we will be able to fix these drawbacks by comparing similar poses in different cycles.



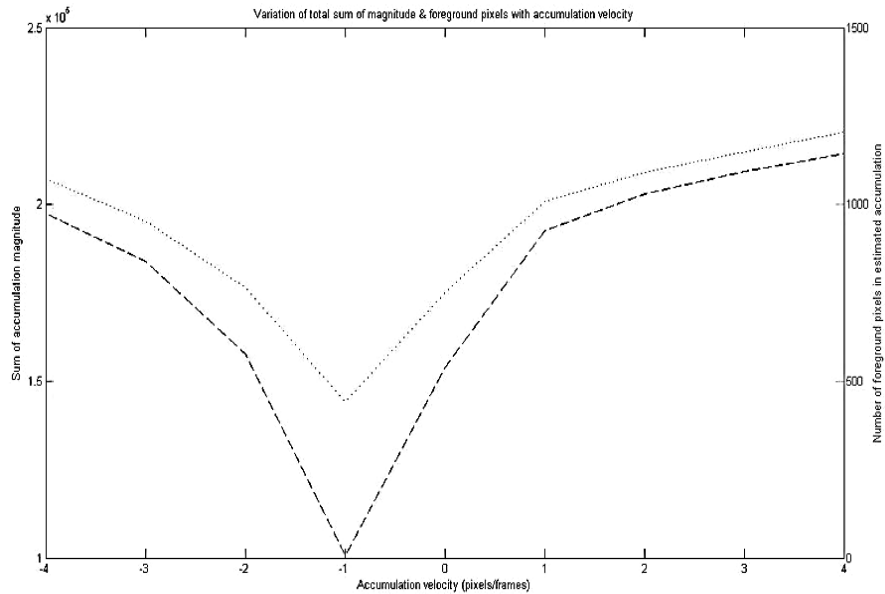
**Fig. 5.2 Output of Foreground Extraction and Moving Edges Detection.**

In the next phase we tried to estimate the speed of walking for each sequence which will be later employed for normalization. Besides, we have to know the direction of walking to be able to compare gait patterns in next steps. This will be

done using the value calculated for speed which declares direction from left to right for positive values of velocity ( $v$ ) and the opposite direction for negative ones. In Fig. 5.3 we have displayed the accumulation for each three different velocities. Comparing the figures we can clearly see that best structured accumulation is obtained for velocity of value -1. This structure gets blurred by increasing or even decreasing this value. Moreover, we can see that those parts of the body which are more stable are more distinguishable in the accumulation. To extract the best velocity, the global maximum of  $\max(A_v)$  which is the graph of maximum intensity value for a range of velocities belongs to the best matched speed value. If none of the objects move in the speed specified in the graph the accumulation would not be dense and thus resulting in a small intensity value. The mentioned graph for the sample displayed in Fig. 5.3 is shown in Fig. 5.4. Two graphs are used to compare the results of checking the sum of intensities (lower graph) and number of pixels (upper graph), which we can see both have the same pattern and global minimum in -1.



**Fig. 5.3 Bulk Motion of Silhouette for a Sample Sequence for Different Velocity Values.**

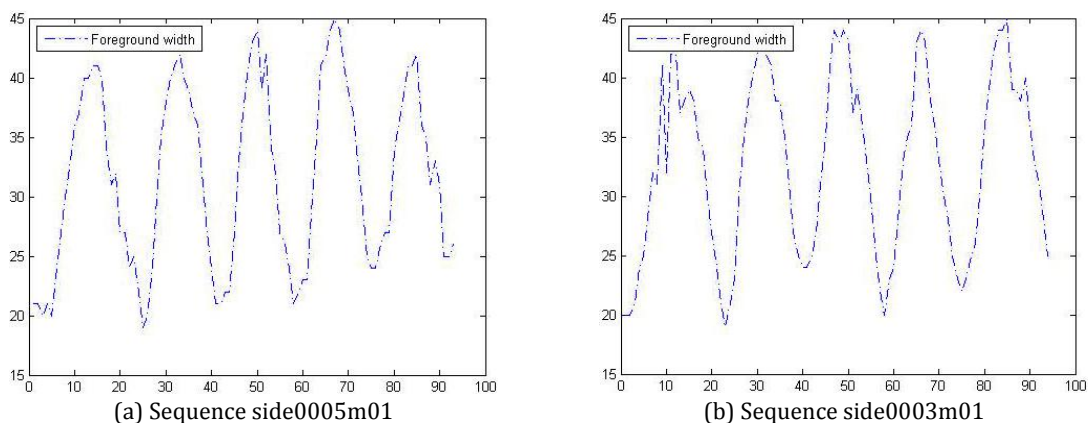


**Fig. 5.4 Variation of Total Intensity based on Velocity in Silhouette Accumulation.**

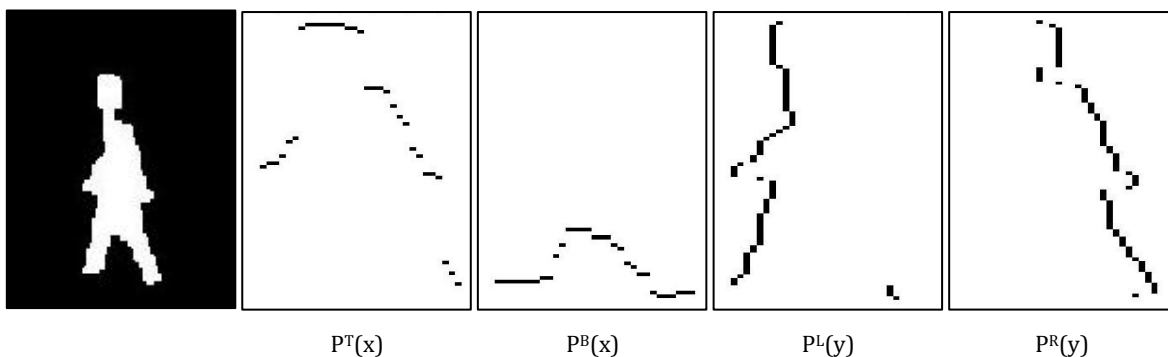
To estimate the period of walking, we have observed the variation of subject's silhouette. To achieve this goal we have used the width of the bounding box which is the smallest rectangle we can fit over the silhouette. This rectangle, thus, has the minimum width where feet are beside each other and the maximum width happens in case of double stance. Therefore, since a cycle includes two steps, in the periodic signal created this way, a cycle would be considered the distance between the current peak to the next one while passing two local minimums in between. The period can be calculated by averaging these values through the whole sequence. Samples of the output signal for two sequences are shown in Fig. 5.5. Considering the periodic nature of walking we can only analyze a cycle of each sequence having calculated the period.

The next step is to create the gait patterns using the model-free algorithm described in section 4.2. We should first build four projections for each frame of

sequence for all the subjects, the output of which is presented in Fig. 5.6 for a sample frame. Having spatial projections we will be able to put them all together and construct the spatio-temporal gait representation images (Fig. 5.7). We can observe that the distance vector is roughly periodic and gives the extent of movement of different parts of the subject; besides, the brighter a pixel in 2D pattern, the larger the value of the distance vector in that position. As stated earlier, training gait patterns are computed from one cycle of each sequence.



**Fig. 5.5 Periodic Pattern of Silhouette Width Variation for Two Sample Sequences.**

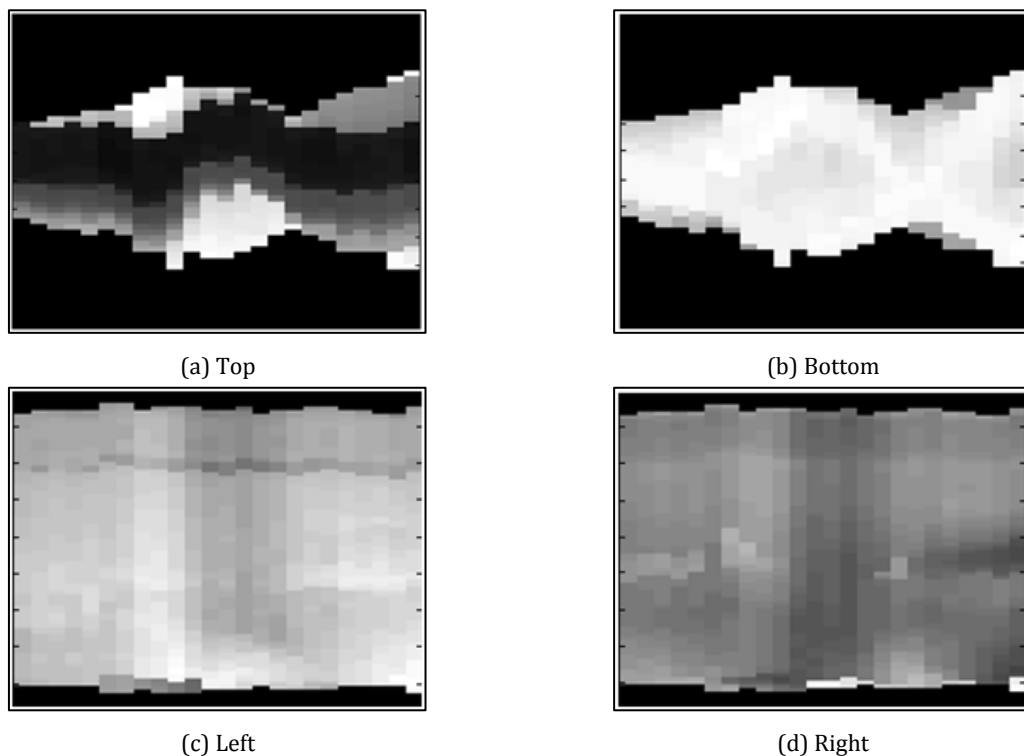


**Fig. 5.6 A Sample Silhouette and its Four Projections for a Selected Frame of Sequence side0001m03.**

To extract the features applying KPCA on temporal contour projection images, a fixed number of frames should have been selected from each sequence to have chromosomes with same length eventually. This value was determined based on the

velocity and period of walking for each individual so that at least the selected frames would include one cycle of walking; the frames also were chosen based on the starting phase of each person. Since every sequence includes more than one cycle, the first cycle has been chosen for each subject to maintain consistency. KPCA was then applied on the temporal contour projections of these frames.

This process led into  $56 \times 30$  eigenvectors for top and bottom gait patterns, and  $86 \times 30$  ones for left and right projections, resulting in 8520 eigenvectors totally. These values were calculated based on a normalized bounding box of  $86 \times 56$  and 30 frames of each sequence.



**Fig. 5.7 Spatio-temporal Gait Patterns estimated for Top, Bottom, Left and Right Projections for the Sequence side0001m03.**

These features were then presented to a GA with population size of 800, 1000 and 1500, during 50, 100 and 20 generations, averaged over 2, 3 and 10 iterations

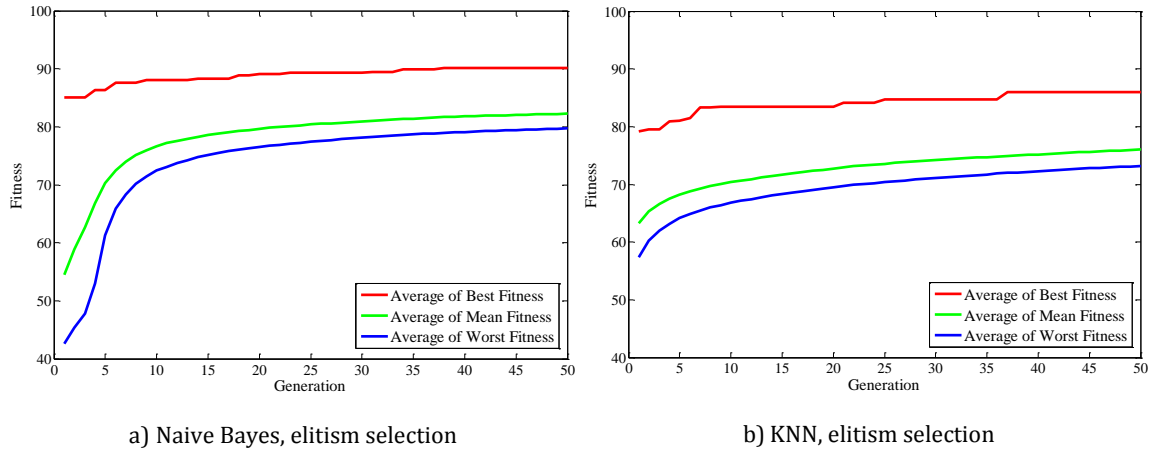
respectively, to observe the effect of various parameters of GA on algorithm performance. The probability of the operators has been chosen to be 0.6 and 0.1 for crossover and mutation respectively in initial experiments. These values were modified later to 0.9 and 0.05 which resulted in better outcome. The fitness function for GA is of the following form with values of 100 and 10 for  $W_A$  and  $W_F$  respectively.

$$\text{Fitness} = 100 \times \text{Correct Recognition Rate (CRR)} - 10 \times \frac{\text{Number of selected features}}{\text{Chromosome length}}$$

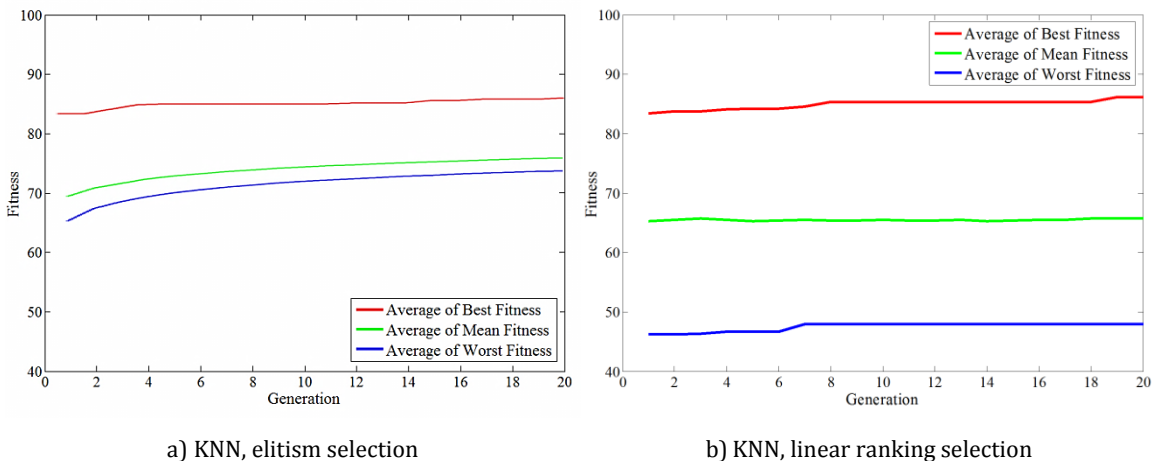
The GA is applied only at the training stage of the system. This reduces resources and time complexity. In addition, the selected subset is minimized and it gets high prediction rates. As mentioned earlier, each subject has 6 sequences, 3 of which are selected randomly as test set. Among the rest of streams, two are employed for training and the remaining one for validation. This process is repeated 7 times. The final CRR is the average value of these iterations. The results are shown in Figures 5.8, 5.9 and 5.10 for different parameter sets.

One important parameter is the size of population; since the length of each chromosome is too long, number of chromosomes in a population should be large enough to be able to contain various types of features permutation in each generation. This approach helps the GA to explore among enough instances and find the individuals with higher discriminatory for next generations. This idea has initially caused the graphs to start from a better starting point in case of having a larger population. The number of generations, on the other hand, should be large enough to let better chromosomes overcome the population gradually. That is why the figures in Fig. 5.9 have not reached an optimal recognition rate. They obviously

need more evolution. Thus, we have increased the number of generations while adding more exploration pressure in the experiments shown in Fig. 5.10.



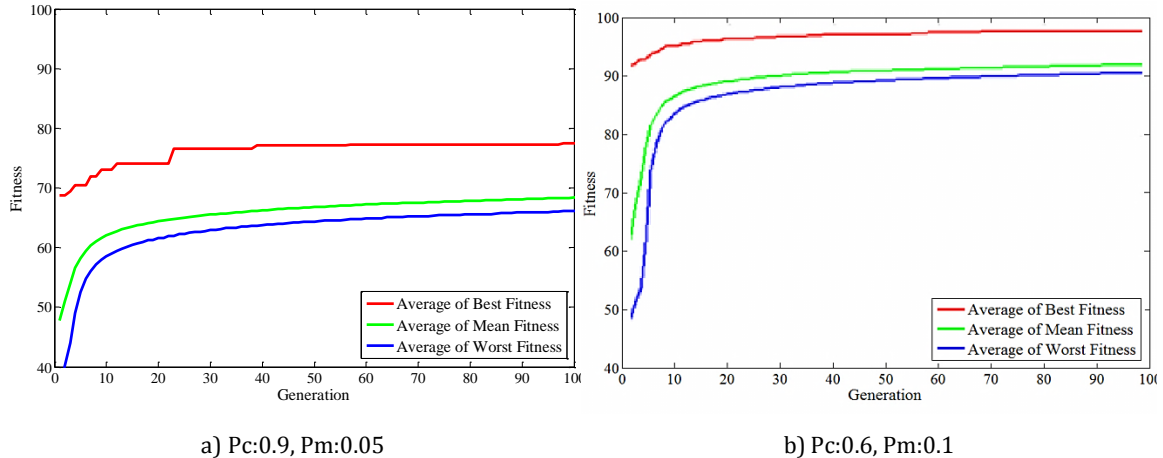
**Fig. 5.8 Results of GA subset selection, population size: 800, generation size: 50. Pc: 0.6, Pm: 0.1, 7-fold cross validation, Comparing the results of different classifiers.**



**Fig. 5.9 Results of GA subset selection, population size: 1500, generation size: 20. Pc: 0.6, Pm: 0.1, 7-fold cross validation, Comparing the results of different selection algorithms.**

The other parameter is the type of classifier which affects the overall fitness directly. Two classifiers have been employed. The first one is Naive Bayes classifier which estimates for every class and every feature separately. Total class densities are constructed by assuming independency and consequently multiplying the separate feature densities. The used default version divides each axis into 10 bins,

counts the number of training examples for each of the classes in each of the bins, and classifies the object to the class that gives maximum posterior probability. Missing values will be put into a separate bin. As explained earlier, Naive Bayes classification is based on estimating the probability or probability density of features  $X$  given class  $Y$ . Here a Gaussian distribution is estimated for each class. The other classifier is KNN which has an estimator with high resolution in regions where the training set is dense. Therefore, the balance between resolution and variance can be adjusted locally. Here a value of 3 has been used for  $K$ .



**Fig. 5.10 Results of GA subset selection, population size: 1000, generation size: 100, Naive Bayes, elitism selection, 7-fold cross validation, Comparing the results of different mutation and recombination probability values.**

Since the evaluation value of each chromosome is determined mainly from the accuracy of classifier, different techniques result in different outputs of features. Here, Naive Bayes has led to a better CCR than the populations evaluated using KNN. This issue, however, needs more analysis for a variety of GA parameters.

The last variable is the type of selection method. Here two selection techniques have been employed, elitism and linear ranking, both of which have got almost the same results in case of parameter variation.

Table 5.1 shows our results before applying GA and Table 5.2 summarizes Recognition rates after feature subset selection with GA using different parameters. According to the above figures and the following table, it is observable that GA in all cases has decreased the number of eigenvectors used in classification (out of 8520) while improving CCR considerably.

**Table 5.1 Averaged Performance in all 6 Experiments before GA**

CCR using all Features	CCR using top 10 Eigenvectors
62.3%	68.8%

**Table 5.2 Overall Performance after GA**

No. of Population	No. of Generation	Classifier	Selection	Probability of Recombination	Probability of Mutation	CCR after GA	Selected Features Percentage
800	50	Naive Bayes	Elitism	0.6	0.1	90.6%	3.1%
800	50	KNN	Elitism	0.6	0.1	88.7%	40.3%
1000	100	Naive Bayes	Elitism	0.9	0.05	78.1%	29.8%
1000	100	Naive Bayes	Elitism	0.6	0.1	<b>96.3%</b>	10.4%
1500	20	KNN	Elitism	0.6	0.1	87.2%	46.0%
1500	20	KNN	Ranking	0.6	0.1	88.0%	22.3%

We have compared the results with two other experiments to have a better overview of the effect of GA on best feature vector subset selection. In experiment 1, the results of which are displayed in first column of Table 5.1, all features have been exploited in classification. In experiment 2, on the other hand, just 10 percent of eigenvectors with higher eigenvalues have been selected as final features which are

supposed to have good discriminatory ability based on the main concept of KPCA (Second column). There is however, considerable improvement in CCR values in comparison with values of 7th column of Table 5.2, exhibiting the effect of GA.

Thus, we can see that the eigenvector subsets selected by the GA approach have improved detection performance. Feature subsets selected by GA yielded an average error rate of 11.85%, better than the 37.7% obtained using the whole feature set or 31.52% using a percentage of the top eigenvectors. These experiments also demonstrate that eigenvectors encoding irrelevant or redundant information have not been favored by the GA approach.

The results of classification show an improved performance compared to the existing approaches. Actually, fair comparisons are made if the implementation conditions and the employed databases are identical. The approach presented in [8] uses the GT database and the static body and stride parameters of subjects. In that work, two sets of parameters are presented and the within and between discrimination power of each set are analyzed. They report an average CCR of 92.5% for the indoor side view images where the people are far and close to the camera. The approach introduced in [10] uses only the trajectories of the lower body joint angles in the GT database. This method has reached a CCR of 73% on 106 time-normalized signals. The approach has gained accurate results through the use of markers, but the classification rate is a demonstration of the deficiency of features. Another method proposed in [7] classifies instances in the same database being able to reach a recognition rate of 91.4% using a model-based approach. We have shown

that the proposed approach has been able to reach a CCR of 96.3%. This recognition rate is really promising in comparison with the methods mentioned above. The above comparisons with the proposed algorithm are summarized in Table 5.3. As mentioned earlier, we are using the indoor sequences as our database as do the approaches exemplified above. Although it might seem a limitation to the proposed algorithm, an approach with capability of utilization in real world applications requires a more sophisticated background subtraction algorithm. Here, we focused on the effect of feature subset selection on a gait recognition technique. However, this issue does not restrict the method and it can be simply generalized to more common situations.

**Table 5.3 Comparison of proposed algorithm with some other approaches**

Approach	# Subjects	Static Parameters	Dynamic Parameters	Algorithm	CCR (%)
[8]	18	Four distances measured at double support phase	---	Silhouette segmentation and body part labeling	92.5
[10]	18	---	Hip and knee joint-angle trajectories	Sensor markers	73.0
[7]	18	Characteristics of mean shape model	Movements of joint positions of leg and arm	Active contour model and Hough transform	94.5
Proposed Alg.	18	Velocity, approximate height and period	---	Using optimal subset of KPCA features of spatio-temporal projections of silhouette using GA	96.3

## CHAPTER 6

### CONCLUSION AND FUTURE WORK

We have proposed a method for extracting the gait signatures and kinematic features for analyzing and identifying the gait motion on which we wanted to prove the ability of GA-based feature selection to improve the performance.

Temporal contour projections have been employed as gait patterns from which we were able to extract required feature vector for classification. First, some static parameters were estimated; these values were later used for data normalization of gait patterns. These gait patterns are in essence spatio-temporal gait representation which are multi-projections of silhouettes for individual recognition by gait. Kernel PCA based feature extraction approach for gait recognition was then presented. Eigenvectors calculated from applying KPCA on four gait patterns of each subject yields a large number of features that can be reduced with GA. The classification results demonstrated the power of GA in selecting the best and smallest subset of feature which led to a promising recognition rate compared to the whole feature vector which we got after KPCA and perform well in identification. The proposed approach achieved highly competitive performance with respect to the published major gait recognition approaches.

In our future work, we will generalize the method to other imaging viewpoints and outdoor data, and further analyze the above features for constructing a method capable of operating in real-world applications.

## BIBLIOGRAPHY

- [1] Z. Zhang, M. Hu, and Y. Wang, "A Survey of Advances in Biometric Gait Recognition", CCB'11 Proceedings of the 6th Chinese conference on Biometric recognition, pp.150-158, December 2011.
- [2] J.Yoo and M. S. Nixon, "Automated Marker-less Analysis of Human Gait Motion for Recognition and Classification", ETRI Journal, Vol. 33, No. 2, pp. 259-266, April 2011.
- [3] M. Okumura, H. Iwama, Y. Makihara, and Yasushi Yagi, "Performance Evaluation of Vision-based Gait Recognition using a Very Large-scale Gait Database", Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), Washington D.C., USA, pp. 1-6, September 2010.
- [4] S. M. Hosseini, A. Nasrabadi, P. Nouri and H. Farsi, "A Novel Human Gait Recognition System", International Journal of Computer and Electrical Engineering, Vol. 2, No.6, pp. 1043-1047, December 2010.
- [5] G. Johansson, 1973. "Visual perception of biological motion and a model for its analysis", PerceptionPsychophys.14, pp. 201-211.
- [6] G. Johansson, "Visual motion perception", Scientific American, Vol. 232, pp. 76-88, June 1975.
- [7] F. Tafazzoli, R. Safabakhsh, "Model-based human gait recognition using leg and arm movements", Engineering Applications of Artificial Intelligence, Vol. 23, No. 8, pp. 1237-1246, December 2010.
- [8] A.F Bobick., A.Y., Johnson, "Gait recognition using static activity-specific parameters", Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), Kauai, Hawaii, pp. 423-430, December 2001.
- [9] Y. Jang-Hee, H. Doosung, M. Ki-Young, and M. S. Nixon, "Automated Human Recognition by Gait using Neural Network", First Workshops on Image Processing Theory, Tools and Applications (IPTA), pp. 1-6, November 2008.
- [10] R. Tanawongsuwan, A.F. Bobick, "Gait recognition from time-normalized joint-angle trajectories in the walking plane", IEEE Computer Vision and Pattern Recognition Conference (CVPR 2001), Kauai, Hawaii, pp. 726-731, December 2001.
- [11] A.H. Wang, J.W. Liu, "A gait recognition method based on positioning human body joints", International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR '07), Beijing, pp. 1067-1071, November 2007.
- [12] D. Cunado, M. S. Nixon, J. N. Carter, "Automatic extraction and description of human gait models for recognition purposes", Computer Vision and Image Understanding, Vol. 90, No. 1, pp. 1-41, March 2003.

- [13] R. Urtasun, P. Fua, "3D Tracking for Gait Characterization and Recognition ", Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 17-22, May 2004.
- [14] S.L. Dockstader, M.J. Berg, A.M. Tekalp, "Stochastic Kinematic Modeling and Feature Extraction for Gait Analysis", IEEE Transactions on Image Processing, Vol. 12, No. 8, pp. 962-976, August 2003.
- [15] F. Jean, A. B. Albu, and R. Bergevin, "Towards view-invariant gait modeling: Computing view-normalized body part trajectories", Pattern Recognition, Vol. 42, No. 11, pp. 2936-2949, November 2009.
- [16] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette Analysis-Based Gait Recognition for Human Identification", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 12, pp. 1505-1518, December 2003.
- [17] N. V. Boulgouris and Z. X. Chi, "Human gait recognition based on matching of body components", Pattern Recognition, Vol. 40, No. 6, pp. 1763-1770, June 2007.
- [18] J. Gu, X. Ding, S. Wang, and Y. Wu, "Action and gait recognition from recovered 3-d human joints", IEEE Transactions on System, Man and Cybernetics, Part B: Cybernetics, Vol. 40, No. 4, pp. 1021-1033, August 2010.
- [19] X. Zhang and G. Fan, "Dual gait generative models for human motion estimation from a single camera", IEEE Transactions on System, Man and Cybernetics, Part B: Cybernetics, Vol. 40, No. 4, pp. 1034-1049, August 2010.
- [20] G. Trivino, A. Alvarez-Alvarez, and G. Bailadorb, "Application of the computational theory of perceptions to human gait pattern recognition", Pattern Recognition, Vol. 43, No. 7, pp. 2572-2581, July 2010.
- [21] J. Wang, M. She, S. Nahavandi and A. Kouzani, "A review of vision-based gait recognition methods for human identification", Digital Image Computing: Techniques and Application (DICTA), IEEE, Piscataway, N.J., pp. 320- 327, December 2010.
- [22] L. Jianyi and Z. Nanning, "Gait History Image: A Novel Temporal Template for Gait Recognition", IEEE International Conference on Multimedia and Expo, pp. 663-666, July 2007.
- [23] P.J. Phillips, S. Sarkar, Z. Liu, I.R. Vega, P. Grother, K. Bowyer, "The humanID gait challenge problem: data sets, performance, and analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 2, pp. 162-177, February 2005.
- [24] R.T. Collins, R. Gross, J. Shi, "Silhouette-based human identification from body shape and gait", Fifth IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA, pp. 366-371, May 2002.
- [25] L. Lee and W. Grimson, "Gait Analysis for Recognition and Classification", Fifth IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA, pp. 148-155, May 2002.

- [26] I. Venkat and P. DeWilde, "Robust gait recognition by learning and exploiting sub-gait characteristics", *International Journal of Computer Vision*, Vol. 91, No. 1, pp. 7-23, January 2011.
- [27] L. Xuelong, S. J. Maybank, Y. Shuicheng, T. Dacheng, and X. Dong, "Gait Components and Their Application to Gender Recognition", *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, Vol. 38, No. 2, pp. 145-155, March 2008.
- [28] J. Han and B. Bhanu, "Individual recognition using gait energy image", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 2, pp. 316-322, February 2006.
- [29] 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, No. 3, pp. 257-267, March 2001.
- [30] 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, No. 11, pp. 977-984, August 2009.
- [31] 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, No. 8, pp. 2904-2910, August 2010.
- [32] M. Hu, Y. Wang, Z. Zhang, and Y. Wang, "Combining spatial and temporal information for gait based gender classification", *20th International Conference on Pattern Recognition (ICPR)*, pp. 3679-3682, August 2010.
- [33] K. Bashir, T. Xiang, and S. Gong, "Gait representation using flow fields", *British Machine Vision Conference*, pp. 1-11, September 2009.
- [34] C. BenAbdelkader, R. Cutler, L. Davis, "Gait recognition using image self-similarity", *EURASIP Journal on Applied Signal Processing*, Vol. 2004, No. 4, pp. 572-585, April 2004.
- [35] L. Wang, T. Tan, W. Hu, and H. Ning, "Automatic gait recognition based on statistical shape analysis", *IEEE Transactions on Image Processing*, Vol. 12, No. 9, pp. 1120-1131, September 2003.
- [36] N. V. Boulgouris and Z. X. Chi, "Gait Recognition Using Radon Transform and Linear Discriminant Analysis", *IEEE Transactions on Image Processing*, Vol. 16, No. 3, pp. 731-740, March 2007.
- [37] V. Kellokumpu, G. Zhao, S. Z. Li, and M. Pietikainen, "Dynamic texture based gait recognition", *Third International Conference on Advances in Biometrics (ICB '09)*, pp. 1000-1009, June 2009.
- [38] Y. Ran, Q. Zheng, R. Chellappa, and T. M. Strat, "Applications of a simple characterization of human gait in surveillance", *IEEE Transactions System, Man and Cybernetics. Part B: Cybernetics*, Vol. 40, No. 4, pp. 1009-1020, August 2010.

- [39] C.H. Chen, J.M. Liang, H. Zhao, "Factorial HMM and Parallel HMM for Gait Recognition", IEEE Transactions on Systems Man and Cybernetics Part C: Applications and Reviews, Vol. 39, No. 1, pp. 114-123, January 2009.
- [40] J. Sun, G. Bebis, R. Miller, "Object detection using feature subset selection", ELSEVIER Pattern Recognition, Vol. 37, No. 11, pp. 2165-2176, November 2004.
- [41] B. Guo and M. S. Nixon, "Gait Feature Subset Selection by Mutual Information ", IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, Vol. 39, No. 1, pp. 36-46, January 2009.
- [42] L.S. Oliveir, N. Benahmed, R. Sabourin, F. Bortolozzi, C.Y. Suen, "Feature subset selection using genetic algorithms for handwritten digit recognition", XIV Brazilian Symposium on Computer Graphics and Image Processing, Florianopolis , Brazil, pp. 362-369, October 2001.
- [43] Z. Sun, G. Bebis, R. Miller, "Boosting object detection using feature selection", IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 290-296, July 2003.
- [44] Z. Sun, G. Bebis, X. Yuan, S. J. Louis, "Genetic Feature Subset Selection for Gender Classification: A Comparison Study", Sixth IEEE Workshop on Applications of Computer Vision, (WACV 2002), pp. 165-170, 2002.
- [45] Y. Chtioui, D. Bertrand, D. Barba, "Feature Selection by a Genetic Algorithm. Application to Seed Discrimination by artificial Vision", Journal of the Science of Food and Agriculture, Vol. 76, No. 1, pp. 77-86, January 1998.
- [46] B. Bhanu, Y. Lin, "Genetic algorithm based feature selection for target detection in SAR images", ELSEVIER Image Vision Computing, Vol. 21, No. 7, pp. 591-608, July 2003.
- [47] R. Collins, Y. Liu, "On-line selection of discriminative tracking features", Ninth IEEE International Conference on Computer Vision, Nice, France, pp. 346-352, October 2003.
- [48] Y. Liu, J. Kender, "Video frame categorization using sort-merge feature selection", IEEE Workshop on Motion and Video Computing, pp. 72-77, December 2002.
- [49] G. Veres, L. Gordon, J. N. Carter, and M. Nixon, "What image information is important in silhouette-based gait recognition?", Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 776-782, July 2004.
- [50] R. K. Begg, M. Palaniswami, and B. Owen, "Support vector machines for automated gait classification", IEEE Transactions on Biomedical Engineering, Vol. 52, No. 5, pp. 828-838, May 2005.
- [51] K. Begg. "A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data", Journal of Biomechanics, Vol. 38, No. 6, pp. 401-408, 2005.

- [52] K. Bashir, T. Xiang and S. Gong, "Feature selection on Gait Energy Image for human identification", IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, pp. 985-988, March 2008.
- [53] J.R. Quinlan, 1986, "Induction of decision trees, Machine Learning 1", pp. 81-106.
- [54] J.R. Quinlan, 1993, "C4.5: Programs for Machine Learning", Morgan Kaufmann, San Francisco.
- [55] E.P. Ephzibah, "Cost Effective Approach on Feature Selection using Genetic and Fuzzy Logic for Diabetes Diagnosis", International Journal on Soft Computing (IJSC), Vol.2, No.1, pp. 1-10, February 2011.
- [56] M. Dash, H. Liu and H. Motoda, "Consistency Based Feature Selection", Fourth Pacific Asia Conference on Knowledge Discovery and Data Mining, Springer-Verlag, pp. 98-109, 2000.
- [57] J. H. Holland, 1992, "Adaption in natural and artificial Systems", MIT Press Cambridge, MA, USA.
- [58] C. Huang and C. Wang, "A GA-based feature selection and parameters optimization for support vector machines", Expert Systems with Applications, Vol. 31, No. 2, pp. 231-240, August 2006.
- [59] H. Vafaie, K. De Jong, "Genetic Algorithms as a Tool for Feature Selection in Machine Learning", Fourth International Conference on Tools with Artificial Intelligence (TAI '92), Arlington, VA, USA, pp. 200-203, November 1992.
- [60] W. Siedlecki, J. Sklansky, "A note on genetic algorithm for large-scale feature selection", Pattern Recognition Letters, Vol. 10, No. 5, pp. 335-347, November 1989.
- [61] Raymer, M. L., W. F. Punch, E. D. Goodman, L. A. Kuhn and A. K. Jain, "Dimensionality Reduction Using Genetic Algorithms", IEEE Transactions on Evolutionary Computation, Vol. 4, pp. 164-171, July 2000.
- [62] J. Yang, V. Honavar, "Feature subset selection using a genetic algorithm", IEEE Intelligent systems, Vol. 13, No. 2, pp. 44-49, 1998.
- [63] Tsai, C. Y., "An iterative feature reduction algorithm for probabilistic neural networks", Omega, Vol. 28, pp. 513-524, 2000.
- [64] W. Jiang, G.H. Er, Q.H. Dai and J.W. Gu, "Similarity-based online feature selection in content-based image retrieval", IEEE Transactions on Image Processing, Vol. 15, No. 3, pp. 702-712, March 2006.
- [65] M. Dash, H. Liu, "Consistency-based search in feature selection", Artificial Intelligence, Vol. 151, No. 1-2, pp. 155-176, June 2003.
- [66] Y. Chtioui, D. Bertrand, D. Barba, "Feature selection by a genetic algorithm, application to seed discrimination by artificial vision", J. Sci. Food Agric. Vol. 76, pp. 77-86, 1998.

- [67] H. Vafaie, I. Imam, "Feature selection methods: genetic algorithms vs. greedy-like search", International Conference on Fuzzy and Intelligent Control Systems, Louisville, KY, March 1994.
- [68] R. Zhang, Ch. Vogler, D. Metaxas, "Human gait recognition at Sagittal plane", Image and Vision Computing, Vol. 25, No. 3, pp. 321-330, March 2007.
- [69] H. Takeda, S. Farsiu, P. Milanfar, "Kernel regression for image processing and reconstruction", IEEE Transactions on Image Processing, Vol. 16, No. 2, pp. 349-366, February 2007.
- [70] D. K. Wagg, M. S. Nixon, "Model-based gait enrolment in real-world imagery", Workshop on Multimodal User Authentication, Santa Barbara, CA, USA, pp. 11-12, December 2003.
- [71] N. R. Howe, A. Deschamps, "Better foreground segmentation through graph cuts", Computer Vision and Pattern Recognition, Vol. 94, No. 11, pp. 49-52, September 2007.
- [72] M. Ekinci, M. Aykut, "Human Gait Recognition based on Kernel PCA Using Projections", Journal of Computer Science and Technology, Vol. 22, No. 6, pp. 867-876, November 2007.
- [73] T.J. Chin and D. Suter, "Incremental Kernel Principal Component Analysis", IEEE Transactions on Image Processing, Vol. 16, No. 6, pp. 1662-1674, June 2007.
- [74] A. L.I. Oliveira, P. L. Braga, R. M.F. Lima and M. L. Cornélio, "GA-based method for feature selection and parameters optimization for machine learning regression applied to software effort estimation", ELSIEVER Information and Software Technology, Vol. 52, No. 11, pp. 1155-1166, July 2010.
- [75] J. Lu, T. Zhao, and Y. Zhang, "Feature selection based-on genetic algorithm for image annotation", ELSIEVER Knowledge-Based Systems, Vol. 21, No. 8, pp. 887-891, December 2008.