Journal of Electronic Imaging 24(1), 013036 (Jan/Feb 2015)

Genetic feature selection for gait recognition

Faezeh Tafazzoli,[a] George Bebis,[a,*] Sushil Louis,[a] and Muhammad Hussain[b]
[a]University of Nevada, Department of Computer Science and Engineering, Reno, Nevada, United States
[b]King Saud University, College of Computer and Information Sciences, Computer Science Department, Riyadh 11543, Saudi Arabia

Abstract. Many research studies have demonstrated that gait can serve as a useful biometric modality for human identification at a distance. Traditional gait recognition systems, however, have mostly been evaluated without explicitly considering the most relevant gait features, which might have compromised performance. We investigate the problem of selecting a subset of the most relevant gait features for improving gait recognition performance. This is achieved by discarding redundant and irrelevant gait features while preserving the most informative ones. Motivated by our previous work on feature subset selection using genetic algorithms (GAs), we propose using GAs to select an optimal subset of gait features. First, features are extracted using kernel principal component analysis (KPCA) on spatiotemporal projections of gait silhouettes. Then, GA is applied to select a subset of eigenvectors in KPCA space that best represents a subject's identity. Each gait pattern is then represented by projecting it only on the eigenvectors selected by the GA. To evaluate the effectiveness of the selected features, we have experimented with two different classifiers: k nearest-neighbor and Naïve Bayes classifier. We report considerable gait recognition performance improvements on the Georgia Tech and CASIA databases. *© 2015 SPIE and IS&T* [DOI: 10.1117/1.JEI.24.1.013036]

Keywords: gait recognition; feature selection; genetic algorithms; kernel principal component analysis.

Paper 13557 received Oct. 3, 2013; accepted for publication Feb. 3, 2015; published online Feb. 25, 2015.

1 Introduction

Robust and reliable means of automatic human identification for surveillance and access control are in great demand today. Biometric-based identification using physiological or behavioral characteristics is of particular importance due to their universality and uniqueness.[1] Many biometric-based authentication methods have been proposed using a wide variety of cues, such as fingerprint, hand, iris, face, and gait. Among them, gait identification (i.e., identifying individuals by the way they walk), has attracted considerable attention due to its ability to ascertain somebody's identity at a distance while being noninvasive and nonperceivable.[2,3] Moreover, gait can be detected and measured at low resolution, which could be valuable when high-resolution face or iris information is not available. There is much evidence from psychophysical, medical, and biomechanical experiments indicating that gait patterns are unique to each individual.[4] Compared with other biometric modalities, gait is also less likely to be obscured.[5] However, human gait analysis involves challenging issues due to the highly flexible structure and self-occlusion of the human body. These issues mandate using complicated processes for the measurement and analysis of gait in markerless sequences.[2] For instance, footwear, physical conditions, such as pregnancy, leg or foot injuries, or even drunkenness can change the manner of walking. Like most biometrics, gait will instinctively change with age.[4]

A variety of gait recognition systems have been proposed in the literature (e.g., see Refs. 1 and 6 for a recent review). Typically, a large number of features are extracted to avoid loss of important gait information. Section 2 presents a review of gait recognition methods with emphasis on gait feature extraction. Features extracted from segmented video sequences, for example, are usually very highly dimensional to better account for gait variations due to different factors (e.g., health, age, body size, weight, speed, etc.) as well as limited understanding of what features might be more appropriate to the underlying gait recognition mechanism.[7] Without employing some kind of feature selection strategy, however, many of the gait features being extracted could be redundant or irrelevant to the gait recognition task. In general, feature selection could provide valuable clues in terms of understanding the underlying distinctness among human gait patterns.

Generally speaking, the choice of features to represent the patterns affects several aspects of the recognition problem. Although many methods have presented notable recognition rates using their own feature vector, it has been observed that beyond a certain point, the inclusion of additional features leads to a worse rather than better performance. This apparent paradox presents us with a feature subset selection problem in the automated design of pattern classifiers. Such a problem refers to the task of identifying and selecting a useful subset of features to be used to represent patterns from a larger set of often mutually redundant or even irrelevant features. Therefore, the main goal of feature subset selection 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 multicriterion optimization function (e.g., number of features and accuracy of classification). Automatic feature subset selection distinguishes the proposed classification method from all other reported approaches. In particular, genetic algorithms (GAs) offer a particularly attractive approach for this kind of problem since they are generally quite effective for rapid global

---

*Address all correspondence to: George Bebis, E-mail: bebis@cse.unr.edu search of large, nonlinear and poorly understood spaces. Moreover, GAs are very effective in solving large-scale problems.[8] Examples include face detection,[9,10] gender classification,[11,12] vehicle detection,[9,10] seed discrimination,[13] target detection,[14] tracking,[15] and video categorization.[16]

The use of feature selection, however, has not been given enough consideration in gait recognition. Some approaches[2,17,18] have mainly considered conventional dimensionality reduction or statistical tools, such as principal component analysis (PCA) and analysis of variance. In Ref. 7, Guo and Nixon used mutual information to measure the utility of selected features in recognition. Begg et al.[19] employed a hill-climbing feature selection algorithm. In another work,[20] a forward feature selection algorithm was utilized. Bashir et al.[21] developed a cross-validation-based approach and an effective measurement of the relevance of gait energy image (GEI) features in their approach.

Motivated by our previous work on gender/face/vehicle classification using feature subset selection,[9,10,12] we propose using GAs for feature selection to improve gait recognition performance. 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 using kernel PCA (KPCA) with the goal of selecting a subset of eigenvectors encoding important information about the target concept of interest. KPCA is chosen to handle highly complex and nonlinear data distributions, and produces nonlinear subspaces for better feature extraction. GAs, then, are used to select those eigenvectors encoding the identity of the subject. This is in contrast to the typical strategy of picking a percentage of the top eigenvectors to represent the target concept, independent of the classification task. Our experimental results show an enhanced correct classification rate (CCR) improvement compared to the results of KPCA feature vectors. This implies that GAs can select eigenvectors encoding mostly information related directly to identity.

The rest of this paper is organized as follows. Section 2 presents a review of gait recognition approaches with emphasis on gait feature extraction. Section 3 briefly reviews the problem of feature selection. Section 4 overviews the proposed approach. Experimental results on different parameters of GA along with a comparison with a complete set of features are presented in Sec. 5. Finally, Sec. 6 concludes the paper and gives some suggestions to improve the current approach in the future.

2 Gait Recognition Review

Gait recognition methods can be broadly divided into two categories: model-based and model-free. In this section, we review representative methods from each category with emphasis on feature extraction.

Model-based methods attempt to explicitly model the human body or motion by employing static and dynamic body parameters, which are typically view and scale invariant. Usually, these methods perform model matching in each frame of a walking sequence so that kinematic parameters, such as trajectories, can be measured. Bobick and Johnson[22] formed two groups of static body parameters by calculating four distances between different human body parts. Yoo and Nixon[2] constructed a two-dimensional (2-D) stick figure by extracting nine coordinates from human body contours based on human anatomical knowledge. In a related work,[11] a model was developed by extracting leg angles based on regression analysis of contour data. Tanawongsuwan and Bobick[23] focused on trajectories of joint angles from marker-based motion capture data, the amplitude-frequency and phase-frequency of which were chosen as gait features. Wang and Liu[24] presented a method based on positioning body joints. Cunado et al.[25] modeled the thigh as a rotatable line for temporal evidence gathering by Hough transform and analyzed the data via discrete Fourier transform. Urtasun and Fua[26] proposed three-dimensional (3-D) temporal models to track and recover motion parameters. Dockstader et al.[27] proposed a hierarchical model, which used a set of thick lines joined at a single point to represent the legs and a periodic pendulum motion model to describe the gait pattern. In Wang et al.'s work,[28] the human body was modeled as 14 rigid parts connected to one another at the joints with a total of 48 degrees of freedom (DOFs). Joint-angle trajectory signals were considered as gait dynamics.[6] Boulgouris and Chi[29] separated the human body into different components and combined the results to form a common distance metric to choose the most contributing parts. Li et al.[30] did similar research by dividing the average silhouettes over a gait cycle into seven different parts. Jean et al.[31] proposed an efficient and promising feature based on the trajectories of the head and feet using optical flow. Gu et al.[32] proposed a method to automatically extract multiple configurations and movement features from the sequence of label-free 3-D volume data. They employed a hidden Markov model (HMM) and an exemplar-based HMM to model these features, respectively. Zhang and Fan[33] presented an approach that included two generative models representing the kinematics and appearances of a gait by a few latent variables. Trivinoa et al.[34] divided a gait cycle into four approximately equal phases. Based on the computational theory of perceptions, the relationships among horizontal acceleration, vertical acceleration, and other indicators were learned by rule-based approach.

Model-free approaches usually employ either shape of binary silhouettes or the whole motion of the walking person's body rather than modeling the whole human body or any part of it. These approaches are insensitive to the quality of silhouettes and have the advantage of low computational costs compared to model-based approaches. However, they are usually not robust to viewpoint change and scale.[6] Collins et al.[35] established a method based on template matching of body silhouettes in key frames. In Ref. 36, Lee and Grimson described a moment-based representation of gait appearance. A baseline algorithm for human identification using spatiotemporal correlation of silhouette images was proposed in Ref. 37 by Sarkar et al. Han and Bhanu[17] employed the idea of a motion-energy image and proposed the GEI for individual recognition. Liu and Zheng[38] developed the gait history image (GHI) to retain temporal information as well as spatial information. Chen et al.[39] proposed the frame difference energy image based on GEI and GHI to address the problem of silhouette incompleteness. Xue et al.[40] applied the wavelet decomposition of GEI to infrared gait recognition due to its robustness to the covariates of holding a ball and loading packages. In Ref. 28, Wang et al. proposed a recognition algorithm using the temporal pattern of gait and employing the distance between pixels along the contour and the shape centroid, on which PCA was performed. Dadashi et al.[41] applied wavelet transform to similar temporal patterns. Hu et al.[42] adopted Gabor filters to decompose the body shape into local orientations and scales, and obtained low-dimensional discriminative representation through the agency of PCA and maximization of mutual information. Venkat and DeWilde[43] divided the averaged silhouette into several overlapped parts. They trained a Bayesian network to evaluate the impact of these parts on identification. Bashir et al.[21] divided the flow field into four parts in accordance with the direction and symbol, and used the weighted sum of these parts for gait recognition.

Some other algorithms pay attention to analyzing the whole shape of silhouettes. BenAbdelkader et al.[44] employed silhouette self-similarity. Boulgouris and Chi[45] applied the Radon transform on binary silhouettes to get a template from gait sequences. Linear discriminate analysis and subspace projection were used to extract Radon template coefficients to construct the feature vector. Kellokumpu et al.[46] considered the accumulation of gait sequence as a XYT 3-D space. They employed 3-D local binary features for histogram extraction. Ran et al.[47] proposed a periodic pattern called double helical signature (DHS), which decomposed a video sequence into slices along the X-t dimension and generated DHS by an iterative local curve embedding algorithm. It was used for segmentation and labeling of body parts in cluttered scenes and load-carrying conditions. In Ref. 48, Chen et al. employed the factorial HMM as a feature-level fusion scheme to fuse different gait features, which was compared with the parallel HMM decision-level fusion scheme.

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 d features, the problem is selecting a subset of size m 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.[9] 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,[49] C4.5, and CART, are some of the most successful supervised learning algorithms. Narendra and Fukunaga presented a Branch and Bound algorithm.[50] A well-known algorithm that relies on relevance evaluation is RELIEF. Subset search algorithms[51] search and capture the goodness of each subset. There are again many algorithms that employ exhaustive, heuristic, and random searches. Clustering algorithms are also used for the feature selection process, a couple examples of which are ROCK and CACTUS.[50]

Generally, feature subset selection algorithms can be classified into two categories based on whether feature selection is performed independent of the learning algorithm used to construct the verifier. If so, 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 inductive and 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.[8] 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.[13]

An alternative is the use of GAs for function optimization. They provide the possibility of exploring high-dimensional spaces about which little is known *a priori* and encoding important information about the target concept of interest in an efficient way. GAs—a form of inductive learning strategy—are adaptive search techniques that have demonstrated substantial improvement over a variety of random and local search methods. This is accomplished by their ability to exploit accumulating information about an initially unknown search space in order to bias a subsequent search into promising subspaces. The major reason for GA's 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.[52]

Siedlecki and Sklansky[53] presented one of the earliest studies of GA-based feature selection in the context of k nearest-neighbor (kNN) classifiers. Yang and Honavar[54] proposed a feature selection approach using GA with rank-based selection strategy and NN for classification. They tested their methods using several benchmark real-world pattern classification problems and reported improved results. Chtioui et al.[55] investigated a GA approach for feature selection in a seed discrimination problem. Vafaie and Imam[56] conducted a comparison between important score, a greedy-like feature selection method, and GA. Using several real-world problems, they found that GAs are more robust at the expense of more computational effort.

4 Method Overview

The technique presented in this paper chooses the best feature vector with the 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 a person accumulated over the whole sequence. Figure 1 shows the overall framework of the approach. First, some preprocessing is applied to the input sequence, including noise reduction, background subtraction, and the formation of the silhouette's global temporal accumulation. These steps result in static parameters, such as approximate height of subject, his/her velocity and period of walking, based on which the pattern will be normalized in the next steps. The spatiotemporal information is represented in a single 2-D gait template by using multiprojection of silhouettes. Feature vectors are constructed using eigenvectors derived by applying KPCA on these gait patterns to project

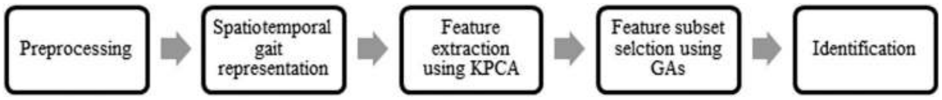


Fig. 1 Proposed algorithm framework.

the data in a lower-dimensional space. The next step performs feature selection using GAs, with the goal of choosing a subset of these eigenvectors, encoding mostly important information about the target concept of interest. The evaluation function used here contains two terms, the first based on classification accuracy on a validation set and the second on the number of eigenvectors selected. Finally, *k*NN and Naïve Bayes classifiers are used to identify subjects in the test set.

Based on physiological research results, most of the walking dynamics take place in the sagittal plane, or the plane that vertically bisects the human body.[1] Thus, our method assumes a single subject moving at a constant speed, fronto-parallel against a static background.

4.1 *Gait Pattern Extraction and Representation*

4.1.1 *Preprocessing*

In this phase, a kernel regression method[57] is first applied on frames of each sequence to remove noise and retain the most representative information of each video corrupted by limitations of the imaging system. The data samples are used to create the initial (dense) estimate of the interpolated output image. In the next iteration, the reconstructed (less noisy) image is used to calculate a more reliable estimate of the gradient, and this process continues for a few more iterations.

This process is followed by background subtraction. The silhouettes extracted in this process will be later used in the stage of gait pattern representation. The method used here for foreground segmentation is based on using the minimum graph-cut method proposed in Ref. 58. In this approach, a graph is built based upon the image in which a typical vertex links to exactly six other nodes: the foreground and the background, plus the vertices of its four-connected neighbors. The weights of the links derive directly from the differences measured between the current frame and the background model at the corresponding pixel. Once constructed, standard methods based upon graph flow will find an optimal cut separating the source from the sink. To build the background model, the pixel color from every fourth frame is taken and the data above and below a pair of thresholds are thrown away. From the remaining numbers, it estimates the mean and variance of each pixel's color, assuming a normal distribution. Segmenting the graph using a standard graph-cut algorithm provides better results compared to morphological-based approaches. They are more successful in overcoming the effects of noise by aggregating information from a local neighborhood around each pixel while remaining true to the underlying data. A typical result of applying foreground extraction is shown in Fig. 2 for two input frame instances.

In the next step, a velocity filtering algorithm is employed to determine the bulk motion of the silhouette of the subject. This algorithm will be applied on the 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. Using a velocity filtering algorithm,[59] it is possible to determine object motion independent of shape, based on

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

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. This algorithm sorts the objects in the scene according to their velocity and starting position, producing an accumulation for each possible object velocity (Fig. 3). The highest peak in the plot of maximal intensity versus velocity indicates the object's velocity.

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. Samples of the output signals are shown in Fig. 4. Considering the periodic nature of walking, we can only analyze a cycle of each sequence having calculated the period.


(a) (b)

Fig. 2 Results of foreground segmentation and preprocessing on sample sequences from (a) Georgia Tech (GT) and (b) CASIA datasets.


v=0 v=-1 v=-2

Fig. 3 Bulk motion of silhouette for a sample sequence for different velocity values.

4.1.2 *Gait representation*

In this paper, we have used an improved spatiotemporal gait representation as the basic gait pattern proposed in Ref. 60. This pattern is obtained by using multiprojections of silhouette. Input silhouettes are first normalized based on the person's approximate height gained through global temporal accumulation of previous phase and then they are horizontally aligned.

Gait pattern is created from the projections of silhouettes (Fig. 5), which are generated from a sequence of binary silhouette images, $B_t(x,y)$, 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 the silhouette. This definition includes four different images for top, bottom, left, and right distance vectors. The values in the top and bottom contour projection vectors are the number of rows between the bounding box and silhouette at each column, which makes their width the same as the width of the bounding box. The same holds for left and right distance vectors, and gives them the same height as the bounding box.

Each gait pattern, hence, is created as a new 2-D image. For instance, a 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 (row sum) of silhouette image $B_t(x,y)$. Thus, $P^T(x,t)$ refers to the count of the number of rows between the top side of the bounding box and the outer contour in the column x of the silhouette image $B_t(x,y)$. The result is a 2-D pattern formed by stacking row projections together to form a spatiotemporal pattern. Other projections, $P^B(x,t)=\sum_{-y}B_t(x,y)$, $P^L(y,t)=\sum_x B_t(x,y)$, and $P^R(y,t)=\sum_{-x}B_t(x,y)$, are constructed in the same way for bottom, left, and right projections. The variation of each component of the distance vectors can be regarded as the gait signature of that object.[60] Figure 6 shows an example of each pattern.

It is clear that the distance vector is roughly periodic and gives the extent of movement of different parts of the subject. The brighter a pixel in the 2-D pattern in the following figure, the larger is the value of the distance vector in that position.

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 the previous step. As a result, training gait features that form the feature database are obtained. This is independently repeated for each gait pattern produced from the projections.

Conventional linear subspace methods, such as 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 data to a higher-dimensional feature space where conventional linear subspace methods can be used, with the resulting subspaces being nonlinear with regard to the original input space. Experiments and comparisons have shown that KPCA almost always outperforms PCA.[61]

To apply KPCA on a data matrix $a=[x_1..x_n]\in R^{m\times n}$, we map it to a higher-dimensional space F using a kernel function $\varphi\colon R^m\to F$ and apply PCA in F. Using φ, a is transformed into $A=[\varphi(x_1)\ldots\varphi(x_n)]$. The map φ is induced by a kernel function $k(.,.)$ that allows efficient evaluation of inner products in $F\colon k(x,z)=\varphi(x)\cdot\varphi(z)$. Considering the matrix $M=\hat{A}^T\hat{A}$ and its eigenvalue

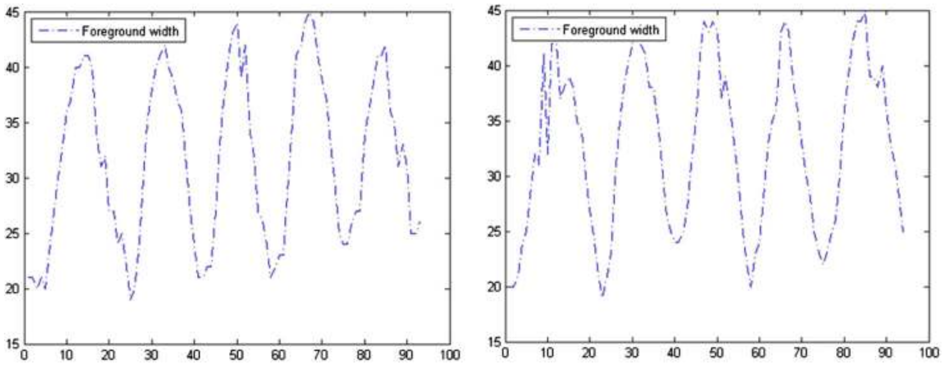


Fig. 4 Periodic pattern of silhouette width variation for two sample sequences.

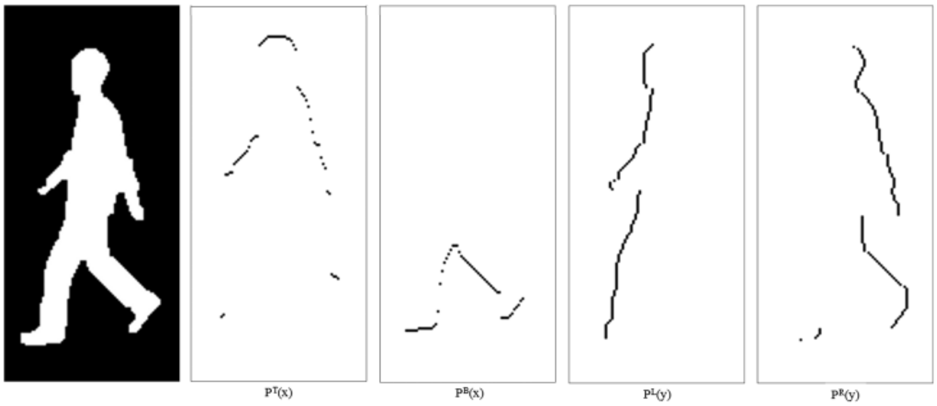


Fig. 5 A sample silhouette and the distance vectors corresponding to its four projections.

decomposition $M = Q\Delta Q^T$, by using $k(.,.)$, A^TA can be evaluated without having to perform the mapping φ since A^TA 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

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

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

$$U^r = A\overset{\prime}{v}Q^r(\Delta^r)^{-1/2} = A\alpha, \tag{3}$$

where $\alpha = \overset{\prime}{v}Q^r(\Delta^r)^{-1/2}$ contains the expansion coefficients. The reconstruction of A using the first-r kernel principal components would be $A^r = \hat{A}^r + \mu_A$.[61]

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 kinds of information.[9] In essence, different tasks make different demands in terms of the information that needs to be processed, and that this information is not contained in the same ranges of eigenvectors.

The number of features that 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, we apply GA to search the space of eigenvectors with the goal of selecting a subset of them which encodes important information about the manner of walking for each individual.

4.3 *Genetic Feature Subset Selection*

Evolutionary algorithms offer a particularly attractive approach to multicriteria optimization because they are effective in high-dimensional search spaces.

A GA is a model of machine learning that 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. 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. 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 the application of genetic operators to generate successive generations of individuals are repeated many times until a termination criterion is satisfied. This termination criterion can be defined as reaching a predefined time limit or number

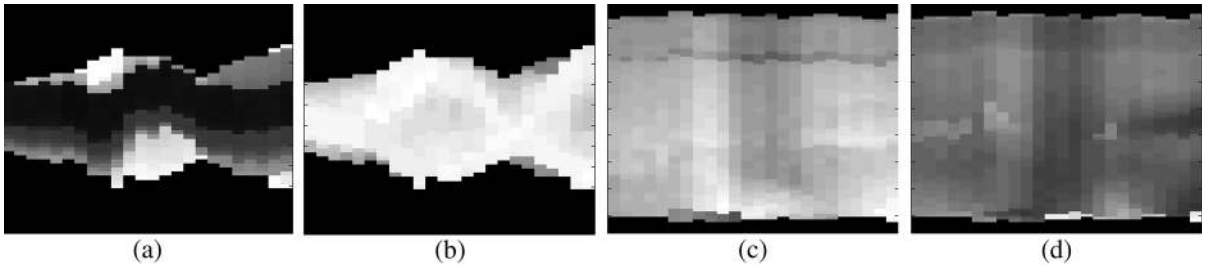

(a) (b) (c) (d)

Fig. 6 Spatiotemporal gait patterns estimated for (a) top, (b) bottom, (c) left, and (d) right projections for a sample sequence from GT.

of generations or population convergence or satisfactory solution. 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 a binary representation. 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.[13]

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 that 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. The overall fitness function will be evaluated by adding the weighted sum of the match score of that bit string indicating present features along with the number of features exploited. However, the former part is the major concern. The final fitness function is defined as

$$F = 100 \times \mathrm{CRR} - \frac{\mathrm{FeatNum}}{\mathrm{ChLen}}, \tag{4}$$

where CRR represents the correct recognition rate and FeatNum is the number of selected features divided by ChLen, the total number of features.

The match score is evaluated based on feedback of some classifiers, which try to classify the input database considering the current available features.

4.3.3 *Initial population*

Generating an initial population of chromosomes is often achieved at random, but the population may be initialized by chromosomes that 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 values are scattered randomly through the whole chromosome. So we would have a population with different permutations and numbers of features. When random initialization of binary chromosomes is used, each bit of the chromosome is randomly set to 0 or 1 according to a probability, which is called the initialization probability.

4.3.4 *Selection*

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 the performance of GA, because it prevents the loss of 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 of being selected. In the proposed method, a uniform crossover is exploited to avoid destroying the schema in case of the presence of dependency among neighbor eigenvectors. Mutation also is applied using the bitwise method. The probability of each of these 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 a better outcome on an exhaustive search basis.

4.4 *Gait Recognition*

Two different classifiers have been utilized to evaluate chromosomes during fitness estimation in this approach, kNN and Naïve Bayes classifiers.

Among the various methods of supervised statistical pattern recognition, the NN rule achieves a consistently high performance without *a priori* assumptions about the distributions from which the training examples are drawn. 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. The kNN classifier extends this idea by taking the k nearest points and assigning the sign of the majority.

The Naïve 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, Naïve Bayes can often outperform more sophisticated classification methods. In simple terms, a Naïve 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 the other features, a Naïve Bayes classifier considers all of these properties to independently contribute to the probability.

5 Experimental Results

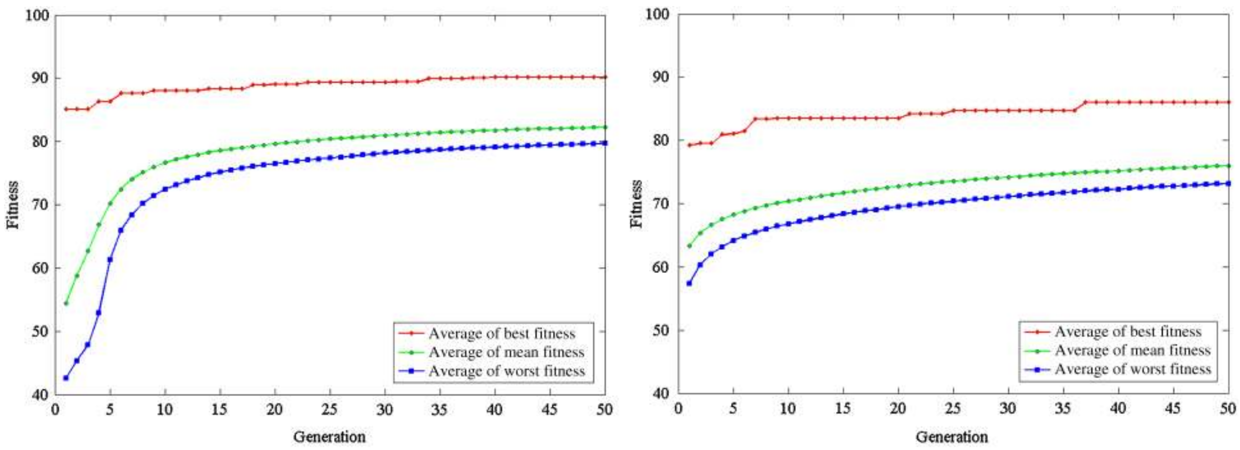
The proposed approach has been tested on Georgia Tech (GT) and CASIA datasets. The GT database[22] consists of 268 sequences from 20 subjects (6 female/14 male). Its sequences were collected under two viewing conditions: side view and angle view (45 deg). For each subject and angle condition, there are six trials. Data collection is done indoors and outdoors. Considering these conditions, we used the indoors data, where lighting is at a constant level and subjects move with approximately constant velocity in front of a plain static background, since the current framework focuses on the feature selection rather than detection and segmentation. 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 (six sequences for 18 subjects). Each video sequence is stored in digital video format, encoded in color PAL format at a resolution of 320 × 240 pixels, and recorded at a rate of 29.97 frames per second (fps). Each sequence typically consists of 80 to 120 frames, or around three full gait cycles.

To better gauge the performance of our method, a second database has been chosen for evaluation. The CASIA dataset B consists of the data from 124 subjects, including 93 males, 31 females, 123 Asians, and 1 European among all subjects. For each subject, six sequences were recorded from his normal walking under different viewing angles in an indoor environment with a simple background.[62,63] The same experiments were repeated for the cases where subjects were wearing a coat or carrying a bag, which were not used in case of our problem. The database consists of many viewing angles ranging from 0 to 180 deg with increments of 18 deg, from which the fronto-parallel sequences where the viewing angle is 90 deg have been utilized. All the video sequences were stored as video files encoded with *mjpeg* codec. The frame size of the video files was 320 × 240, with a frame rate of 25 fps. There were two to three gait cycles in each sequence.[63]

To extract the features and apply KPCA on temporal contour projection images, a fixed number of frames should be selected from each sequence to eventually have chromosomes with the same length. This value was determined based on the velocity and period of walking for each individual so that the selected frames would include at least 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 was chosen for each subject to maintain consistency. KPCA was then applied on the temporal contour projections of these frames using Gaussian function as the kernel. Normalized bounding boxes of size 86 × 56 and 141 × 56 were considered for GT and CASIA datasets, respectively. For each case, a fixed number of frames were chosen based on the starting point of the cycle for all the sequences. For GT 30 and for CASIA 25 frames were the average values including a complete cycle. In the case of GT sequences, the aforementioned process led in to 56 × 30 eigenvectors for top and bottom gait patterns, and 86 × 30 ones for left and right projections, resulting in 8520 eigenvectors total; the same process led to building 9850 eigenvectors for the CASIA video streams.

These features were presented to a GA with a 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 number of generations was later increased for the CASIA dataset to achieve a convergence; these values are included in Table 2.

As mentioned earlier, each subject has six sequences, three of which were selected randomly as a test set. Among the rest of the streams, two were employed for training and the remaining one for validation. This process was repeated seven times. The final CRR is the average value of these iterations. The results are shown in Figs. 7–12 for different parameter sets on both datasets. Each graph includes three plots; best fitness represents the best fitness value achieved in each generation, where mean and worst are the average of all fitness values and the worst one in the



(a) (b)

Fig. 7 Results of GA subset selection on GT dataset, population size: 800, generation size: 50. Pc: 0.6, Pm: 0.1, sevenfold cross-validation, comparing the results of different classifiers. a) Naïve Bayes, Elitism selection; b) KNN, Elitism selection.

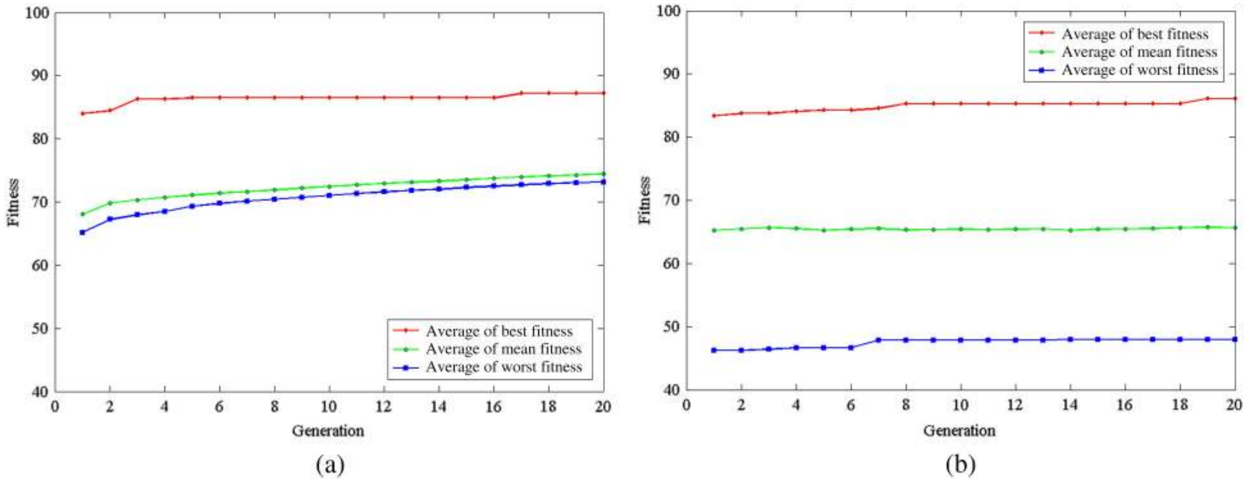


Fig. 8 Results of GA subset selection on GT dataset, population size: 1500, generation size: 20. Pc: 0.6, Pm: 0.1, sevenfold cross-validation, comparing the results of different selection algorithms. (a) KNN, Elitism selection; (b) KNN, Ranking selection.

current generation, respectively. These values are averages over different numbers of iterations for each experiment.

Another important parameter is the size of the population; since the length of each chromosome is too long, the number of the population should be large enough to be able to contain various types of feature permutations in each generation. This approach helps the GA to explore among enough instances and find the individuals with higher discrimination for next generations. This idea has initially caused the graphs to start from a better starting point in the case of having a larger population. The number of generations, on the other hand, should be large enough to let better chromosomes gradually overcome the population. That is why the graphs in Fig. 8 have not reached an optimal recognition rate. They obviously need more evolution. Thus, we increased the number of generations while adding more exploration pressure in the experiments of Fig. 9.

The other parameter is the type of classifier, which directly affects the overall fitness. Two classifiers have been employed. The first one is Naïve 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 were put into a separate bin. Naïve 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 *k*NN, which has an estimator with a 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*. Since the evaluation value of each chromosome is determined mainly from the accuracy of the classifier, different techniques result in different outputs of features. Here, Naïve Bayes has led to a better CCR than the populations evaluated using *k*NN for GT samples. This issue, however, needs more analysis for a variety of GA parameters. Then, we tried adding Support vector machine (SVM) to check whether we are able to improve the results.

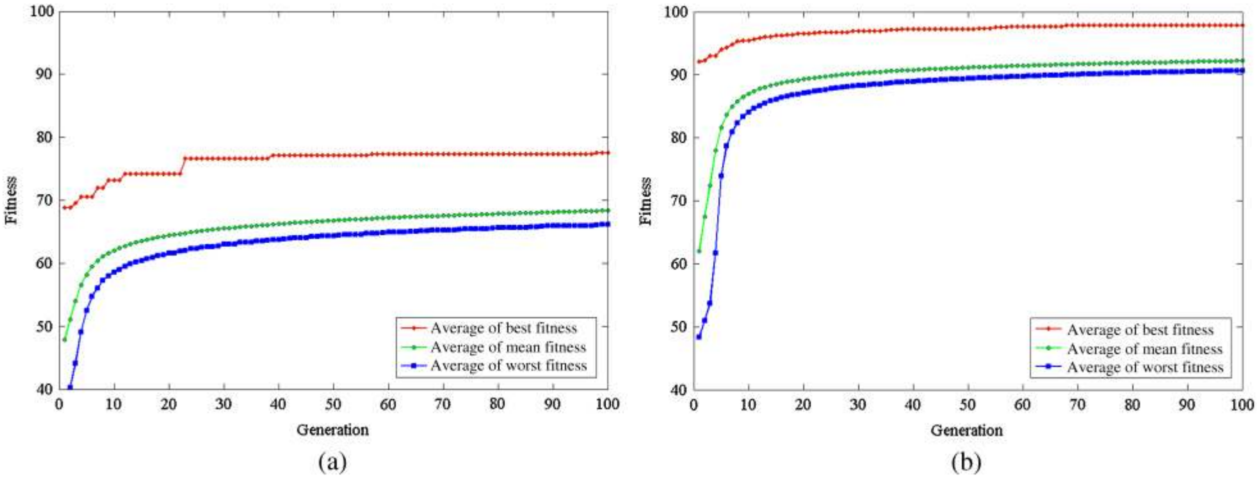
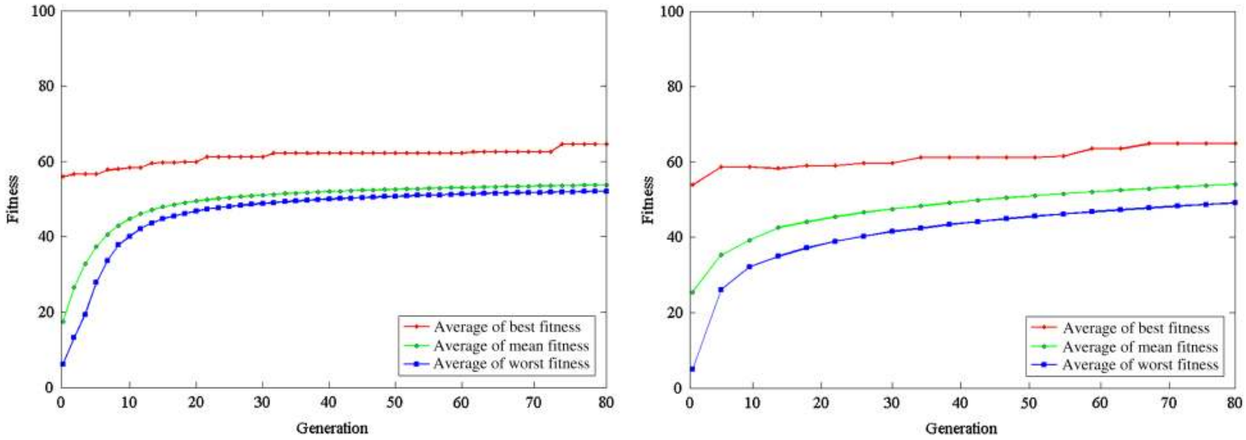


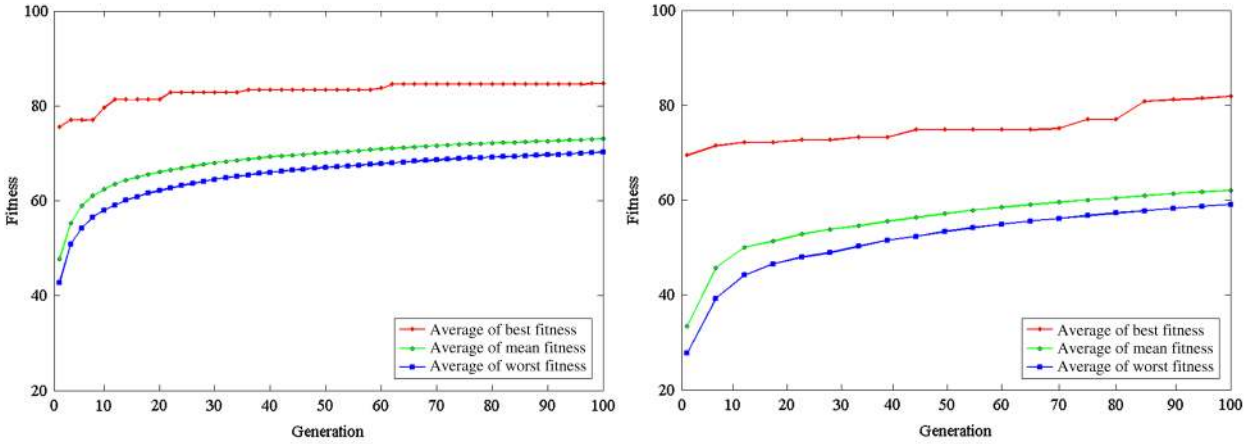
Fig. 9 Results of GA subset selection on GT dataset, population size: 1000, generation size: 100, Naïve Bayes, elitism selection, sevenfold cross-validation, comparing the results of different mutation and recombination probability values. (a) Pc: 0.9, Pm: 0.05; (b) Pc: 0.6, Pm: 0.1.



(a)

(b)

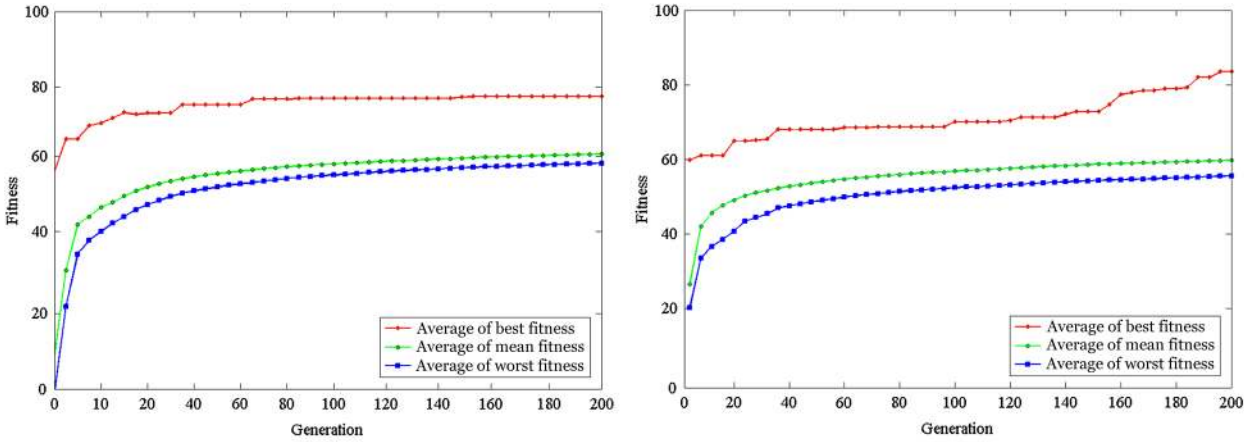
Fig. 10 Results of GA subset selection on CASIA dataset, population size: 800, generation size: 80. Pc: 0.6, Pm: 0.1, sevenfold cross-validation, comparing the results of different classifiers. (a) Naïve Bayes, Elitism selection; (b) KNN, Elitism selection.



(a)

(b)

Fig. 11 Results of GA subset selection on CASIA dataset, population size: 1500, generation size: 100. Pc: 0.6, Pm: 0.1, sevenfold cross-validation, comparing the results of different selection algorithms. (a) KNN, Elitism selection; (b) KNN, Linear Ranking selection.



(a)

(b)

Fig. 12 Results of GA subset selection on CASIA dataset, population size: 1000, generation size: 200, Naïve Bayes, elitism selection, sevenfold cross-validation, comparing the results of different mutation and recombination probability values. (a) Pc: 0.9, Pm: 0.05; (b) Pc: 0.6, Pm: 0.1

Table 1 Averaged performance in all six experiments before genetic algorithm (GA) for Georgia Tech (GT) and CASIA datasets.

Dataset	CCR using all features (%)	CCR using top 10% eigenvectors (%)
GT	62.3	68.8
CASIA	77.7	77.8

CCR, correct classification rate.

As we have shown in Table 2, SVM did not improve the accuracy for the GT dataset. However, we were able to choose a better subset of features with this new classifier for the CASIA dataset.

The last variable is the type of selection method. Here, two selection techniques have been employed, elitism and linear ranking, both of which have obtained almost the same results in the case of parameter variation. The same experiments have been performed for the data from the CASIA database. Interestingly, parameter sets have performed almost the same. However, due to the considerable increase in the variation of data and size, these experiments needed a larger number of generations to converge.

Table 1 shows our results before applying GA, and Table 2 summarizes recognition rates after feature subset selection with GA using different parameters. According to these figures and tables, it is observable that GA in all cases has decreased the number of eigenvectors used in classification (out of 8520 and 9850 for GT and CASIA datasets, respectively) while improving CCR considerably.

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 the first column of Table 1, all features have been exploited in the classification. In experiment 2, on the other hand, just 10% of the eigenvectors with higher eigenvalues have been selected as the final features, which are supposed to have good discriminatory ability based on the main concept of KPCA (third column). There is, however, considerable improvement in CCR values in comparison with values of the seventh column of Table 2 exhibiting the effect of GA. Feature subsets selected by GA for the GT dataset yielded an average error rate of 24.55%, better than the 37.7% obtained using the whole feature set or 31.52% using a percentage of the top eigenvectors. The same process was repeated for the CASIA dataset. The improvement, however, has not been that great since, as it was mentioned before, more generations are required to let GA converge. But the average error rate has still decreased by 20.0%.

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 where the implementation conditions and the databases employed for experiments are identical. The approach presented in Ref. 22 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 powers 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 Ref. 23 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 the features. Another method proposed in Ref. 64 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 3.

As mentioned earlier, we are using the indoor sequences in our database as do the approaches referred to above.

Table 2 Overall performance after GA for GT and CASIA datasets.

# pop	Classifier	Selection	Crossover prob.	Mutation prob.	# gen	CCR after GA for GT (%)	% selected features	# gen	CCR after GA for CASIA (%)	% selected features
800	Naïve Bayes	Elitism	0.6	0.1	50	90.6	3.1	80	65.7	14.7
800	KNN	Elitism	0.6	0.1	50	88.7	40.3	80	63.2	28.5
800	SVM	Elitism	0.6	0.1	50	81.3	41.8	80	73.6	32.5
1000	Naïve Bayes	Elitism	0.9	0.05	100	78.1	29.8	200	79.3	14.4
1000	Naïve Bayes	Elitism	0.6	0.1	100	**96.3**	10.4	200	84.9	19.8
1500	KNN	Elitism	0.6	0.1	20	87.2	46.0	100	86.3	31.8
1500	KNN	Ranking	0.6	0.1	20	88.0	22.3	100	81.8	26.4
1500	SVM	Elitism	0.6	0.1	20	84.6	35.7	100	**88.6**	44.1

KNN, K nearest neighbor.

Table 3 Comparison of proposed algorithm with some other approaches.

Approach	# subjects	Static parameters	Dynamic parameters	Alg.	CCR (%)
Ref. 22	18	Four distances measured at double support phase	—	Silhouette segmentation and body part labeling	92.5
Ref. 23	18	—	Hip and knee joint-angle trajectories	Sensor markers	73.0
Ref. 64	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 kernel principal component analysis features of spatiotemporal projections of silhouette using GA	96.3

Although it might seem a limitation to the proposed algorithm, an approach with the 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.

We have performed another comparison based on the selected features. In this experiment, we chose GEI features, which have proven to achieve good results in the case of gait recognition. Conventional silhouette based gait representations treat gait as a sequence of templates. In contrast, GEI represents gait using a single image, which contains information about both body shape and human walking dynamics. GEI is, thus, a compact representation, which makes it an ideal starting point for feature selection since it is computationally expensive if the number of features to select is high.[62] Table 4 demonstrates the results of recognition using features extracted by KPCA from GEI representations. We can see that GEI features have been completely successful in discriminating between subjects of the GT dataset. On the contrary, their performance on the CASIA data has not been as good as the results of Table 1. The purpose of this experiment was to see whether these features perform better than multiple projection-based features. We should mention that in this experiment only the normal sequences with a view point of 90 deg have been employed to maintain consistency with our previous experiments. One of the reasons that GEI did not perform well in the second dataset is that although it is comparatively robust to noise, it loses the dynamical variation between successive frames, which was emphasized in the case of contour projections. To check the effect of feature selection on a different type of feature, we applied GA with the parameters achieving better results in our previous experiments on this data. The results of this set of experiment have been included in Table 4. As we can see, in the case of the CASIA dataset, the accuracy has increased considerably using appropriate features from selected eigenvectors of GEI images.

Table 4 Recognition results using gait energy image features for GT and CASIA datasets.

Dataset	CCR using all features (%)	CCR using top 10% eigenvectors (%)	CCR using feature subset selected from GA (%)
GT	100	100	100
CASIA	58.8	46.6	82.9

6 Conclusions

We have described a method for extracting the gait signatures and kinematic features for analyzing and identifying the gait motion. Temporal contour projections have been employed as gait patterns from which we were able to extract required the feature vectors for classification. First, some static parameters were estimated; these values were later used for data normalization from which gait patterns were extracted. Eigenvectors calculated by applying KPCA on four gait patterns of each subject yielded a large number of features that can be reduced with GA. The classification results demonstrated the power of GA in selecting the best subset of feature, which led to a promising recognition rate compared to the whole feature vector we obtained with KPCA.

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.

References

1. Z. Zhang, M. Hu, and Y. Wang, “A survey of advances in biometric gait recognition,” in *Proc. of the 6th Chinese Conf. on Biometric Recognition*, pp. 150–158, Springer Berlin Heidelberg, Berlin (2011).
2. J. Yoo and M. S. Nixon, “Automated marker-less analysis of human gait motion for recognition and classification,” *ETRI J.* **33**(2), 259–266 (2011).
3. M. Okumura et al., “Performance evaluation of vision-based gait recognition using a very large-scale gait satabase,” in *Fourth IEEE Int. Conf. on Biometrics: Theory, Applications and Systems*, pp. 1–6, IEEE, Piscataway, New Jersey (2010).
4. G. Johannson, “Visual perception of biological motion and a model for its analysis,” *Percept. Psychophys.* **14**, 201–211 (1973).
5. S. M. Hosseini et al., “A novel human gait recognition system,” *Int. J. Comput. Electr. Eng.* **2**(6), 1043–1047 (2010).
6. J. Wang et al., “A review of vision-based gait recognition methods for human identification,” in *Int. Conf. on Digital Image Computing: Techniques and Application*, pp. 320–327, IEEE, Piscataway, New Jersey (2010).
7. B. Guo and M. S. Nixon, “Gait feature subset selection by mutual information,” *IEEE Trans. Syst. Man Cybern. A Syst. Hum.* **39**(1), 36–46 (2009).

8. L. S. Oliveir et al., "Feature subset selection using genetic algorithms for handwritten digit recognition," in *XIV Brazilian Symp. on Computer Graphics and Image Processing*, pp. 362–369, IEEE, Piscataway, New Jersey (2001).
9. Z. Sun, G. Bebis, and R. Miller, "Object detection using feature subset selection," *Pattern Recognit.* **37**(11), 2165–2176 (2004).
10. Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection using evolutionary Gabor filter optimization," *IEEE Trans. Intell. Transport. Syst.* **6**(2), 125–137 (2005).
11. J. Yoo et al., "Automated human recognition by gait using neural network," in *First Workshops on Image Processing Theory, Tools and Applications*, pp. 1–6, IEEE, Piscataway, New Jersey (2008).
12. Z. Sun et al., "Genetic feature subset selection for gender classification: a comparison study," in *Sixth IEEE Workshop on Applications of Computer Vision*, pp. 165–170, IEEE, Piscataway, New Jersey (2002).
13. Y. Chtioui, D. Bertrand, and D. Barba, "Feature selection by a genetic algorithm, application to seed discrimination by artificial vision," *J. Sci. Food Agric.* **76**(1), 77–86 (1998).
14. B. Bhanu and Y. Lin, "Genetic algorithm based feature selection for target detection in SAR images," *Image Vis. Comput.* **21**(7), 591–608 (2003).
15. R. Collins and Y. Liu, "On-line selection of discriminative tracking features," in *Ninth IEEE Int. Conf. on Computer Vision*, pp. 346–352, IEEE Computer Society, Washington, DC (2003).
16. Y. Liu and J. Kender, "Video frame categorization using sort-merge feature selection," in *IEEE Workshop on Motion and Video Computing*, pp. 72–77, IEEE, Piscataway, New Jersey (2002).
17. J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(2), 316–322 (2006).
18. G. Veres et al., "What image information is important in silhouette-based gait recognition?," in *Proc. of the 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Vol. 2, pp. 776–782, IEEE, Piscataway, New Jersey (2004).
19. R. K. Begg, M. Palaniswami, and B. Owen, "Support vector machines for automated gait classification," *IEEE Trans. Biomed. Eng.* **52**(5), 828–838 (2005).
20. K. Begg, "A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data," *J. Biomech.* **38**, 401–408 (2005).
21. K. Bashir, T. Xiang, and S. Gong, "Gait representation using flow fields," in *British Machine Vision Conf.*, pp. 1–11, The British Machine Vision Association (BMVA) Press, Durham (2009).
22. A. F. Bobick and A. Y. Johnson, "Gait recognition using static activity-specific parameters," in *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 423–430, IEEE, Piscataway, New Jersey (2001).
23. R. Tanawongsuwan and A. F. Bobick, "Gait recognition from time-normalized joint-angle trajectories in the walking plane," in *IEEE Computer Vision and Pattern Recognition Conf.*, pp. 726–731, IEEE, Piscataway, New Jersey (2001).
24. A. H. Wang and J. W. Liu, "A gait recognition method based on positioning human body joints," in *Int. Conf. on Wavelet Analysis and Pattern Recognition*, pp. 1067–1071, IEEE, Piscataway, New Jersey (2007).
25. D. Cunado, M. S. Nixon, and J. N. Carter, "Automatic extraction and description of human gait models for recognition purposes," *Comput. Vis. Image Underst.* **90**(1), 1–41 (2003).
26. R. Urtasun and P. Fua, "3D tracking for gait characterization and recognition," in *Sixth IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 17–22, IEEE, Piscataway, New Jersey (2004).
27. S. L. Dockstader, M. J. Berg, and A. M. Tekalp, "Stochastic kinematic modeling and feature extraction for gait analysis," *IEEE Trans. Image Process.* **12**(8), 962–976 (2003).
28. L. Wang et al., "Silhouette analysis-based gait recognition for human identification," *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(12), 1505–1518 (2003).
29. N. V. Boulgouris and Z. X. Chi, "Human gait recognition based on matching of body components," *Pattern Recognit.* **40**(6), 1763–1770 (2007).
30. X. Li et al., "Gait components and their application to gender recognition," *IEEE Trans. Syst. Man Cybern. C Appl. Rev.* **38**(2), 145–155 (2008).
31. F. Jean, A. B. Albu, and R. Bergevin, "Towards view-invariant gait modeling: computing view-normalized body part trajectories," *Pattern Recognit.* **42**(11), 2936–2949 (2009).
32. J. Gu et al., "Action and gait recognition from recovered 3-D human joints," *IEEE Trans. Syst. Man Cybern. B Cybern.* **40**(4), 1021–1033 (2010).
33. X. Zhang and G. Fan, "Dual gait generative models for human motion estimation from a single camera," *IEEE Trans. Syst. Man Cybern. B Cybern.* **40**(4), 1034–1049 (2010).
34. G. Trivinoa, A. Alvarez-Alvareza, and G. Bailadorb, "Application of the computational theory of perceptions to human gait pattern recognition," *Pattern Recognit.* **43**(7), 2572–2581 (2010).
35. R. T. Collins, R. Gross, and J. Shi, "Silhouette-based human identification from body shape and gait," in *Fifth IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 366–371, IEEE, Piscataway, New Jersey (2002).
36. L. Lee and W. Grimson, "Gait analysis for recognition and classification," in *Fifth IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 148–155, IEEE, Piscataway, New Jersey (2002).
37. S. Sarkar et al., "The humanID gait challenge problem: data sets, performance, and analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(2), 162–177 (2005).
38. J. Liu and N. Zheng, "Gait history image: a novel temporal template for gait recognition," in *IEEE Int. Conf. Multimedia and Expo*, pp. 663–666, IEEE, Piscataway, New Jersey (2007).
39. C. Chen et al., "Frame difference energy image for gait recognition with incomplete silhouettes," *Pattern Recognit. Lett.* **30**(11), 977–984 (2009).
40. Z. Xue et al., "Infrared gait recognition based on wavelet transform and support vector machine," *Pattern Recognit.* **43**(8), 2904–2910 (2010).
41. F. Dadashi, B. N. Araabi, and H. Soltanian-Zadeh, "Gait recognition using wavelet packet silhouette representation and transductive support vector machines," in *2nd Int. Congress on Image and Signal Processing*, pp. 1–5, IEEE, Piscataway, New Jersey (2009).
42. M. Hu et al., "Combining spatial and temporal information for gait based gender classification," in *20th Int. Conf. on Pattern Recognition*, pp. 3679–3682, IEEE, Piscataway, New Jersey (2010).
43. I. Venkat and P. DeWilde, "Robust gait recognition by learning and exploiting sub-gait characteristics," *Int. J. Comput. Vis.* **91**(1), 7–23 (2011).
44. C. BenAbdelkader, R. Cutler, and L. Davis, "Gait recognition using image self-similarity," *EURASIP J. Appl. Signal Process.* **2004**(4), 572–585 (2004).
45. N. V. Boulgouris and Z. X. Chi, "Gait recognition using Radon transform and linear discriminant analysis," *IEEE Trans. Image Process.* **16**(3), 731–740 (2007).
46. V. Kellokumpu et al., "Dynamic texture based gait recognition," in *Third Int. Conf. on Advances in Biometrics*, pp. 1000–1009, Springer Berlin Heidelberg, Berlin (2009).
47. Y. Ran et al., "Applications of a simple characterization of human gait in surveillance," *IEEE Trans. Syst. Man Cybern. B Cybern.* **40**(4), 1009–1020 (2010).
48. C. H. Chen, J. M. Liang, and H. Zhao, "Factorial HMM and parallel HMM for gait recognition," *IEEE Trans. Syst. Man Cybern. C Appl. Rev.* **39**(1), 114–123 (2009).
49. J. R. Quinlan, "Induction of decision trees," *Mach. Learn.* **1**(1), 81–106 (1986).
50. E. P. Ephzibah, "Cost effective approach on feature selection using genetic and fuzzy logic for diabetes diagnosis," *Int. J. Soft Comput.* **2**(1), 1–10 (2011).
51. M. Dash, H. Liu, and H. Motoda, "Consistency based feature selection," in *Fourth Pacific Asia Conf. on Knowledge Discovery and Data Mining* Vol. **1805**, pp. 98–109, Springer Berlin Heidelberg, Berlin (2000).
52. H. Vafaie and K. De Jong, "Genetic algorithms as a tool for feature selection in machine learning," in *Fourth Int. Conf. on Tools with Artificial Intelligence*, pp. 200–203, IEEE, Piscataway, New Jersey (1992).
53. W. Siedlecki and J. Sklansky, "A note on genetic algorithm for large-scale feature selection," *Pattern Recognit. Lett.* **10**(5), 335–347 (1989).
54. J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," in *IEEE, Intelligent Systems and their Applications* Vol. 13(2), 44–49 (1998).
55. Y. Chtioui, D. Bertrand, and D. Barba, "Feature selection by a genetic algorithm, application to seed discrimination by artificial vision," *J. Sci. Food Agric.* **76**, 77–86 (1998).
56. H. Vafaie and I. Imam, "Feature selection methods: genetic algorithms vs. greedy-like search," in *Int. Conf. on Fuzzy and Intelligent Control Systems*, Institution for Fuzzy Systems and Intelligent Control, Louisville, Kentucky (1994).
57. H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. Image Process.* **16**(2), 349–366 (2007).
58. N. R. Howe and A. Deschamps, "Better foreground segmentation through graph cuts," *Comput. Vis. Pattern Recognit.* **94**(11), 49–52 (2004).
59. D. K. Wagg and M. S. Nixon, "Model-based gait enrolment in real-world imagery," in *Workshop on Multimodal User Authentication*, pp. 11–12, UC Discovery Grant (2003).
60. M. Ekinci and M. Aykut, "Human gait recognition based on kernel PCA using projections," *J. Comput. Sci. Technol.* **22**(6), 867–876 (2007).
61. T. J. Chin and D. Suter, "Incremental kernel principal component analysis," *IEEE Trans. Image Process.* **16**(6), 1662–1674 (2007).
62. S. Zheng et al., "Robust view transformation model for gait recognition," in *Int. Conf. on Image Processing*, pp. 2073–2076, IEEE, Piscataway, New Jersey (2011).
63. S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *Proc. of the 18'th Int. Conf. on Pattern Recognition*, pp. 441–444, IEEE, Piscataway, New Jersey (2006).

64. F. Tafazzoli and R. Safabakhsh, "Model-based human gait recognition using leg and arm movements," *Eng. Appl. Artif. Intell.* **23**(8), 1237–1246 (2010).

Faezeh Tafazzoli received her master's degree in computer engineering from Amirkabir University of Technology in 2008. She received her second master's degree in computer science in 2012 from the University of Nevada, Reno (UNR). She is currently pursuing her PhD studies in computer science at the University of Louisville. Her research interests include content-based image/video retrieval and object and motion tracking.

George Bebis received his BS degree in mathematics and his MS degree in computer science from the University of Crete in 1987 and 1991, respectively, and his PhD degree in electrical and computer engineering from University of Central Florida in 1996. Currently, he is professor, director of Computer Vision Laboratory and chair in the Department of Computer Science and Engineering at UNR. His research has been funded by NSF, NASA, ONR, NIJ, and Ford Motor Company.

Sushil Louis is a professor and director of the Evolutionary Computing Systems Laboratory in the Department of Computer Science and Engineering, UNR, Nevada, USA. He received his PhD from Indiana University, Bloomington, in 1993 and is a member of the IEEE and ACM. More information on his current work is available at his website http://www.cse.unr.edu/~sushil.

Muhammad Hussain is an associate professor in the Department of Software Engineering, King Saud University, Saudi Arabia. He received both his MSc and MPhil from University of the Punjab, Lahore, Pakistan, in 1990 and 1993, respectively. In 2003, he received his PhD in computer science, specializing in computer graphics from Kyushu University, Fukuoka, Japan. He worked as a researcher at the Japan Science and Technology Agency from April 2003 to September 2005.