Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Feature selection for face recognition: a memetic algorithmic approach

Dinesh KUMAR[1], Shakti KUMAR[2], C. S. RAI[3]

([1]Department of Computer Science & Engineering, Guru Jambheshwar University of Science & Technology, Haryana 125001, India)

([2]Institute of Science and Technology, Haryana 135001, India)

([3]University School of Information Technology, GGS Indraprastha University, Delhi 110403, India)

E-mail: dinesh_chutani@yahoo.com; shakti@istk.org; csrai_ipu@yahoo.com

Received June 18, 2008; Revision accepted Oct. 23, 2008; Crosschecked June 10, 2009

Abstract: The eigenface method that uses principal component analysis (PCA) has been the standard and popular method used in face recognition. This paper presents a PCA - memetic algorithm (PCA-MA) approach for feature selection. PCA has been extended by MAs where the former was used for feature extraction/dimensionality reduction and the latter exploited for feature selection. Simulations were performed over ORL and YaleB face databases using Euclidean norm as the classifier. It was found that as far as the recognition rate is concerned, PCA-MA completely outperforms the eigenface method. We compared the performance of PCA extended with genetic algorithm (PCA-GA) with our proposed PCA-MA method. The results also clearly established the supremacy of the PCA-MA method over the PCA-GA method. We further extended linear discriminant analysis (LDA) and kernel principal component analysis (KPCA) approaches with the MA and observed significant improvement in recognition rate with fewer features. This paper also compares the performance of PCA-MA, LDA-MA and KPCA-MA approaches.

Key words: Face recognition, Memetic algorithm (MA), Principal component analysis (PCA), Linear discriminant analysis (LDA), Kernel principal component analysis (KPCA), Feature selection

doi:10.1631/jzus.A0820460 **Document code:** A **CLC number:** TP391.4

INTRODUCTION

Face recognition is becoming a very hot and interesting research area because of its potential applications (Chellappa *et al.*, 1995; Zhao *et al.*, 2003), in particular involving the great concern of security and privacy. As an ideal application for security, face recognition can be deployed to limit an employee's access to sensitive data in private companies and to prevent the physicians from accessing their patients' records in hospitals and the others like airport security, criminal identification, video surveillance, law enforcement, etc. Its approaches are broadly grouped into feature-based and appearance-based techniques. In the first case, geometric characteristics (local statistics of the eyes, the nose, the mouth, etc.) of the faces to be matched are compared, whereas in the second case, the faces are represented as a two-dimensional array of pixel intensity values. Appearance-based approach extracts holistic representation of the whole face image and comparison is done with several such face images. The popular recognition techniques most frequently used with face recognition are the eigenface method (Turk and Pentland, 1991), independent component analysis (ICA) (Bartlett *et al.*, 1998; Bartlett, 2001), Fisher's linear discriminant analysis (LDA) (Swets and Weng, 1996; Belhumeur *et al.*, 1997), etc. Kirby and Sirovich (1990) have used for the first time principal component analysis (PCA) to represent images of human faces, and PCA subsequently applied to face recognition by Turk and Pentland (1991) was popularly known as the eigenface method and became one of the standard and most widely used face recognition approaches. The eigenface method linearly projects the image space to a feature space of a lower dimensionality. Another most popular technique for dimensionality reduction in face recognition application is LDA, also known as Fisherfaces. The two methods of PCA and LDA differ in that the former aims to extract a subspace in which the variance is maximized and to seek the directions that are efficient for representation, whereas the latter seeks directions that are efficient for discrimination. PCA deals with second order statistics whereas the high order statistical dependencies remain untouched. The higher order statistical dependencies are very well dealt with by ICA for face recognition. Kernel principal component analysis (KPCA) (Schölkolf *et al.*, 1998; Yang *et al.*, 2000; Kim *et al.*, 2002), as a nonlinear extension of PCA, is another technique that takes into account the higher order statistics. Therein the input space is first mapped into feature space via nonlinear mapping and then principal components are determined in the feature space.

Feature selection becomes increasingly important because of its demands in various science and engineering disciplines such as pattern recognition, biometrics, remote sensing, data mining, and knowledge discovery (Dash and Liu, 1997; Jain and Zongker, 1997; Kohavi and John, 1997). The goal of feature selection is to find a minimal feature subset out of a given feature set where redundant features are removed to reduce feature space while retaining the information and accuracy in representing the original data. Conventional feature selection methods such as correlation coefficients, residual mean square, and mutual information (Holz and Loew, 1994; Kwak and Choi, 2002) are based on the statistical tools, while the other class of feature selection methods includes methods derived from population-based soft computing optimization algorithms. These algorithms render feature selection as an optimization issue and find an optimal feature subset. Evolutionary algorithms (EAs) are stochastic search methods that mimic the metaphor of natural biological evolution and/or the social behaviour of species (Elbeltagi *et al.*, 2005). Genetic algorithms (GAs) as the first evolutionary based techniques were developed based on the Darwinian principle of the 'survival of the fittest' and the natural process of evolution through reproduction (Goldberg, 1989). Thanks to its demonstrated ability to reach near-optimum solutions to large problems, the GA techniques have been used in many applications (Siedlecki and Sklansky, 1989; Merz and Freisleben, 1997; Oh *et al.*, 2004). In an attempt to improve the quality of solutions and particularly to avoid being trapped in local optima, other EAs have also been introduced during the past decade. In addition to various GAs improvements, recent developments in EAs include other techniques inspired by different natural processes: memetic algorithms (MAs) (Moscato, 1999), particle swarm optimization (PSO) (Bonabeau *et al.*, 1999), ant colony systems (Dorigo and Caro, 1999; Dorigo and Stutzle, 2004), etc.

In this paper we propose feature selection for face recognition using MAs. The general idea behind MAs is to combine the advantages of evolutionary operators that determine interesting regions in the search space with local neighbourhood searches that quickly find good solutions in a small region of the search space. Feature selection helps reduce variations of face images and at the same time enhances the discriminating capability. In this proposed method the objective function for feature selection is the recognition rate of the system and Euclidean norm is used as the nearest neighbour classifier. The same approach has also been applied to LDA and KPCA for the comparison of all the three methods, i.e., PCA-MA, LDA-MA and KPCA-MA.

OVERVIEW OF FEATURE SELECTION METHODS

According to the evaluation procedure, the feature selection algorithms (Dash and Liu, 1997; Kohavi and John, 1997; Xing *et al.*, 2001; Guyon *et al.*, 2002; Hsu *et al.*, 2004; Mao, 2004) can be broadly classified into two categories, i.e., the filter method and the wrapper method. With the filter method, feature selection is done independently of the algorithm, making use of intrinsic characteristics of the data to find the feature subset. The wrapper method, on the other hand, performs feature selection by using the learning algorithm to evaluate the feature subset. These two methods are further classified as forward selection, backward elimination, forward/backward combination, random choice, and instance based methods. The search for feature selection may start with no features, all features or a random subset of features, and the features are added successively as the search proceeds in the first case. Search starting with all features will eliminate the features successively while in the last case the features could be added, removed or reproduced randomly. The first two cases bear the disadvantage that once being selected (eliminated) the features cannot be discarded (reselected) later. Thus the floating search method is proposed to empower a flexibility of adding the discarded features or dropping the already selected features (Pudil *et al.*, 1994). Amongst a number of search procedures we can see such popular ones as stepwise search (Kittler, 1978), branch-and-bound (Narendra and Fukunaga, 1977), GAs (Oh *et al.*, 2004), and ant colony optimization (Yan and Yuan, 2004).

In this paper, a novel wrapper based approach has been proposed using a memetic framework through the integration of a GA and local search. A feature subset of a predefined number of features is generated by the generation procedure and then evaluated in comparison with the previous best candidate. The previous subset is replaced if the generated feature subset is better, and the operation of feature selection will be determined by the stopping criterion.

EIGENFACE METHOD FOR FACE RECOGNITION

The eigenface method as one of the most successful appearance-based face recognition approaches adopts PCA. PCA is an unsupervised technique linearly transforming an original set of variables into a substantially smaller set of uncorrelated variables that represent most of the information in the original set of variables. Such a technique transforms the data to a new coordinate system where the largest variance by any projection of the data comes to lie on the first coordinate (this first coordinate is known as the first principal component), the second largest variance on the second coordinate, and so on. These principal components are orthogonal. As the theoretically optimum linear transform for the given data in least mean square errors, PCA is used to reduce the dimensionality of the original high dimensional dataset by compressing them into lower dimensions and reconstructing the original data. A small set of uncorrelated variables is much easier to understand and use in further analysis than a larger set of correlated variables. In 1901, Pearson first introduced this transform in a biological context to recast linear regression analysis into a new form. It was employed by Hotelling (1993) in the context of psychometry to transform discrete variables into uncorrelated coefficients and hence also known as Hotelling transform or Karhunen-Loéve transform in some of its applications. Kirby and Sirovich (1990) were the first to apply PCA to reconstruct human faces, showing that any face can be represented along the eigenpictures coordinate space and faces can also be reconstructed using a small number of eigenpictures and the corresponding coefficients. The technique proposed by Turk and Pentland (1991) gave an idea of eigenfaces. They made the first successful demonstration of machine recognition of faces. Eigenfaces were to be the eigenvectors associated with the corresponding eigenvalues, determined by computing the covariance of training face image data. Only the best eigenvectors with the dominant eigenvalues were retained, resulting in dimensionality reduction and simplification in face recognition process as well.

Consider a set of N sample images $\{\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \ldots, \boldsymbol{\Gamma}_N\}$ taking values in an n-dimensional image space. Assuming each face image has $m\times n=M$ pixels, and is represented as an $M\times 1$ column vector, a ‘training set’ T with N face images of known individuals forms an $M\times N$ matrix:

$$\boldsymbol{\Gamma}=[\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \ldots, \boldsymbol{\Gamma}_N]. \tag{1}$$

$\boldsymbol{\Psi}$ is the mean image of all the samples:

$$\boldsymbol{\Psi}=\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i. \tag{2}$$

Subtracting $\boldsymbol{\Psi}$ from all sample images yields a new set $\boldsymbol{F}=[\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \ldots, \boldsymbol{\Phi}_N]$, where $\boldsymbol{\Phi}_i=\boldsymbol{\Gamma}_i-\boldsymbol{\Psi}$. By applying PCA to the new set $\boldsymbol{F}$, we obtain a set of N orthonormal vectors $\boldsymbol{V}_i$. The kth vector is chosen such that

$$\lambda_k=\sum_{i=1}^{N}(\boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{\Phi}_i)^2 \tag{3}$$

is maximum subject to

$$\boldsymbol{V}_j^{\mathrm{T}}\boldsymbol{V}_k = \begin{cases} 1, & \text{if } j=k, \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

$\boldsymbol{V}_k$ and λ_k are the eigenvectors and eigenvalues, respectively, of the covariance matrix

$$\boldsymbol{C} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Phi}_i\boldsymbol{\Phi}_i^{\mathrm{T}} = \boldsymbol{F}\boldsymbol{F}^{\mathrm{T}}. \quad (5)$$

We only choose d eigenvectors $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, …, $\boldsymbol{V}_d$ of $\boldsymbol{C}$ that correspond to the d largest eigenvalues λ_1, λ_2, …, λ_d. The new image $\boldsymbol{f}_{\text{new}}$ is then projected onto the chosen set of eigenvectors, $\boldsymbol{U}=[\boldsymbol{V}_1, \boldsymbol{V}_2, …, \boldsymbol{V}_d]$, to give (Turk and Pentland, 1991)

$$\omega_k = \boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{f}_{\text{new}} \quad \text{for } k=1, 2, …, d. \quad (6)$$

The ω_k (k=1, 2, …, d) form a weight distribution vector

$$\boldsymbol{\omega}=[\omega_1, \omega_2, …, \omega_d]. \quad (7)$$

Euclidean norm has been used as the measure of similarity. The test image (new image) weight vector is matched with those of known (training) images.

MEMETIC ALGORITHMS

Evolutionary algorithms (EAs), used as a general term for evolutionary programming, evolution strategies, genetic algorithms, and genetic programming, have been applied successfully in various domains of search, optimization, and artificial intelligence (Merz and Freisleben, 2000). In the field of combinatorial optimization, augmenting EAs with problem-specific heuristics has been shown to be capable of yielding highly effective approaches. These hybrid EAs combine the advantages of efficient heuristics by incorporating domain knowledge and population-based search approaches. One form of such hybridization is the use of local search in EAs which, sometimes called genetic local search algorithms, belong to the class of MAs.

MAs are evolutionary algorithms applying a separate local search process to refine individuals to improve their fitness. These are population based metaheuristic search methods inspired by Darwin's principles of natural evolution and rely on the concept of biological evolution and Dawkin's concept of a meme (defined as a unit of mimic cultural evolution that is capable of local refinements) (Dawkins, 1976). While in nature genes are usually not modified during an individual's lifetime, memes are. The unique aspect of MAs is that all chromosomes and offspring are allowed to gain some experience through a local search before being involved in the evolutionary process. By using an EA to perform exploration, MAs combine global and local search and make a powerful algorithmic technique for evolutionary computing while the local search method performs exploitation. Enjoying a wide variety of applications (Aydemir *et al.*, 2005; Sheng *et al.*, 2007; 2008; Zhu and Ong, 2007; Zhu *et al.*, 2007), MAs search more efficiently than their conventional counterparts and converge to high quality solutions. A pseudo code for an MA procedure is given in Fig.1. Similar to the GAs, an initial population is created at random and a local search is performed on each population member afterwards to improve its experience and thus to obtain a population of local optimum solutions. Then, crossover and mutation operators are applied to produce offspring which are henceforth subjected to the local search so that local optimality is always maintained (Elbeltagi *et al.*, 2005).

```
Procedure MA
begin
 generate random population of P solutions (chromosomes);
 for each individual i∈P: calculate fitness(i);
   for j=1 to #generations
     for each individual i∈P: do i=Local-Search(i);
       for crossover
         select two parents i_a, i_b∈P randomly;
         generate offspring i_c=Crossover(i_a, i_b);
         i_c=Local-Search(i_c);
         add individual i_c to P;
       end for
       for mutation
         select an individual i∈P randomly;
         generate offspring i_c=Mutate(i);
         i_c=Local-Search(i_c);
         add individual i_c to P;
       end for
       P=select(P);
       j=j+1;
   end for
 end
```

Fig.1 Pseudo code for a memetic algorithm procedure

MEMETIC ALGORITHM FOR FEATURE SELECTION

Face recognition system

Within a face recognition system based on PCA and MA, PCA is first of all applied to the face image data and the eigenvectors are computed and then arranged in the descending order of their corresponding eigenvalues. Only the significant eigenvectors with the dominant eigenvalues are retained. This is done to reduce the dimensionality. For our work, before feature selection, we retained only thirty percent of the total number of eigenvectors. For the projection of the data, which of these retained eigenvectors are to be used or not, is decided by the MA. After feature (eigenvector) selection, the training data are projected onto these eigenvectors and PCA coefficients are calculated. Face recognition is done using the test images by comparing the coefficients of the test images with coefficients of those present in the training set. Euclidean norm has been used as the distance measure for classification of images. Finally, after a number of generations, the subset of features (eigenvectors) with the highest recognition rate is considered to be the optimal subset.

Objective function

The objective function is defined by the recognition rate, i.e.,

$$Fitness(ch)=F(S_{ch}), \tag{8}$$

where S_{ch} denotes the corresponding subset of a predefined number of selected features in chromosome *ch*. For a given subset the evaluation is done by the feature selection criterion function $F(S_{ch})$. In this paper $F(S_{ch})$ is specified as the recognition rate for the selected feature subset.

Proposed MA for feature selection

Fig.2 shows the framework for feature selection through the MA procedure for face recognition. Subsequent to the random creation of an initial population, a local search is performed on each population member to improve its experience and thus to obtain a population of local optimum solutions. Crossover and mutation operators are afterwards applied to produce offspring, which are then subjected to the local search so that local optimality is always maintained.

The parameters used for the memetic algorithm are presented in Table 1. Fig.3 shows the flowchart for the proposed algorithm.

begin
  define strategy parameters for the memetic algorithm;
  iteration=0;
  create initial population of chromosomes;
  while *iteration*≤maximum iteration
   constrain chromosomes if required;
   get the features (eigenvectors) for each meme in the chromosome;
   evaluate each set of a predefined number of features and calculate the recognition rate using Euclidean norm as the distance measure;
   apply the MA procedure to get the new population;
   iteration=*iteration*+1;
  end while
end

Fig.2 A framework for feature selection through the memetic algorithm

Table 1 Parameters used in the memetic algorithm

Parameter	Value
Population size	15.00
Crossover rate (XOVR)	0.80
Mutation probability (p_m)	0.05
Generation gap (GGAP)	0.80
Local search probability (p_{ls})	0.50

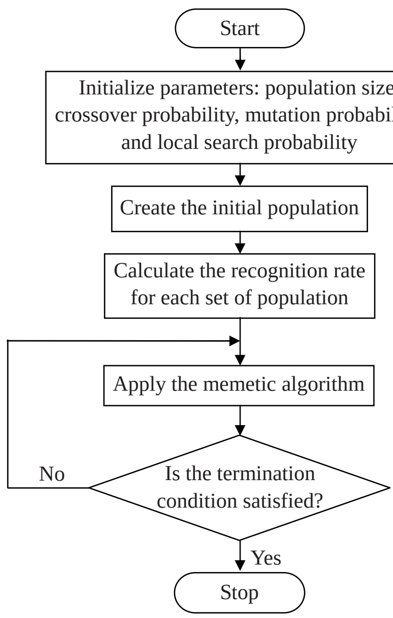


Fig.3 Flow chart for feature selection using the memetic algorithm

By following the above mentioned procedure we obtain the optimal solutions in terms of the maximum recognition rate with a predefined number of selected features, and the selection of features is determined by the MA. The block diagram in Fig.4 represents the complete system.

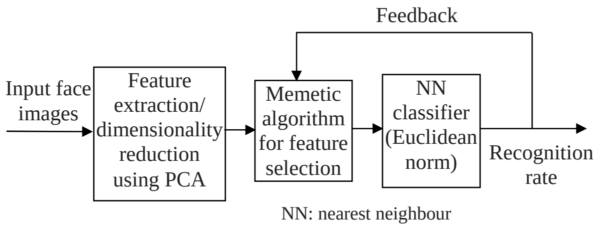


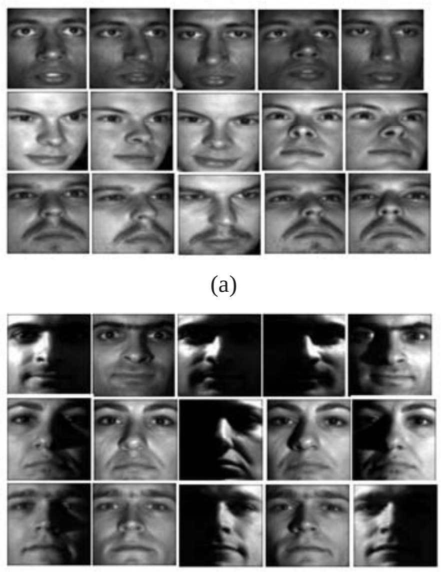
Fig.4 Block diagram for the proposed feature extraction and selection scheme

EXPERIMENTAL RESULTS AND DISCUSSION

Our experimental work deployed two widely used benchmark face image databases, ORL (available at http://www.cam-orl.co.uk/facedatabase.html) and YaleB (Georghiades *et al.*, 2001; Lee *et al.*, 2005) for face recognition. The ORL face database is composed of 400 images with each image having a resolution of 92×112. As many as 40 different persons are contained in the database and each person has his/her 10 different images (for sample images, see Fig.5). These images have been taken at different time, varying in lighting slightly, facial expressions (open/close eyes, smiling/no-smiling), and facial details (glasses/no-glasses). All the images are taken against a dark homogeneous background and the subjects are in up-right, front position with slight left right rotation.

Fig.5 Some images of the ORL database

The extended Yale face database B contains 16128 images of 28 human subjects under 9 poses and 64 illumination conditions. We have considered 10 subjects for experimentation and comparison. The original images were first cropped and then resized to 168×192. Two different sets were prepared. One set contained 10 subjects with each having 9 poses under one lighting condition indicating the pose variations while the other set had 10 subjects with each having the same pose but 10 different illumination conditions. Figs.6a and 6b show some sample images of the two sets of this database. For computational simplicity, the images were resized to 60×60 prior to further processing.



(a)

(b)

Fig.6 Some images of the YaleB database

(a) Each subject has different poses under one lighting condition; (b) Each subject has the same pose but under different illumination conditions

In our experiments, the images for each class were first randomized. Two different cases were considered for the experimental results:

Case 1: We have taken 3 out of a total of 10 images from each class to be used as the training data and kept the remaining 7 (6 for the YaleB Pose set) as the test data.

Case 2: Five images per person were taken for training and the remaining 5 (4 for the YaleB Pose set) as the test data.

Feature extraction and dimensionality reduction

The training data were first histogram equalized. With the application of PCA, the features were extracted in the form of eigenvectors. Only thirty percent of the total number of eigenvectors was retained and the dimensionality was reduced. This was done for both of the cases.

Feature selection

After extracting the eigenvectors, the MA was used to select the optimal feature subset for predefined numbers of features. The MA parameters selected are given in Table 1. In order to obtain the optimal values for the various parameters, the value of mutation probability (p_m) was first varied from 0.002 to 0.02 to 0.05 and the optimal results were finally found at 0.05. The crossover rate and the local search probability were further varied from 0.5 to 0.8 and 0.4 to 0.8, respectively, but no change was observed in the performance. This may be attributed to the smaller search space. The experiments were carried out for 10, 20, 30 and 40 classes for the ORL database.

Experiment 1 This experiment was carried out for Case 1 where 3 images per person were taken for training while the remaining 7 images for testing for different numbers of classes.

Table 2 describes the subsets of features selected that yield the maximum recognition rate for different numbers of classes.

The results (Figs.7a~7d) show that as the number of features increases, the recognition rate of the face

Table 2 Subsets of features selected (Case 1, ORL)

Number of classes	Selected features (eigenvectors)	Number of selected features	Maximum recognition rate (%)
10	(1,2,5,6,7,7,8,9)	8	98.57
20	(1,3,6,7,8,10,11,12,12,13,15,15,16,17)	14	96.43
30	(1,3,4,5,6,7,8,9,10,11,13,14,15,17,18,19,20,21,22,24,25,27)	22	93.33
40	(1,3,4,5,6,7,8,9,10,13,14,15,16,17,18,19,20,21,22,23,25,26, 28,29,30,30,31,32,33,34,35,36)	32	86.79

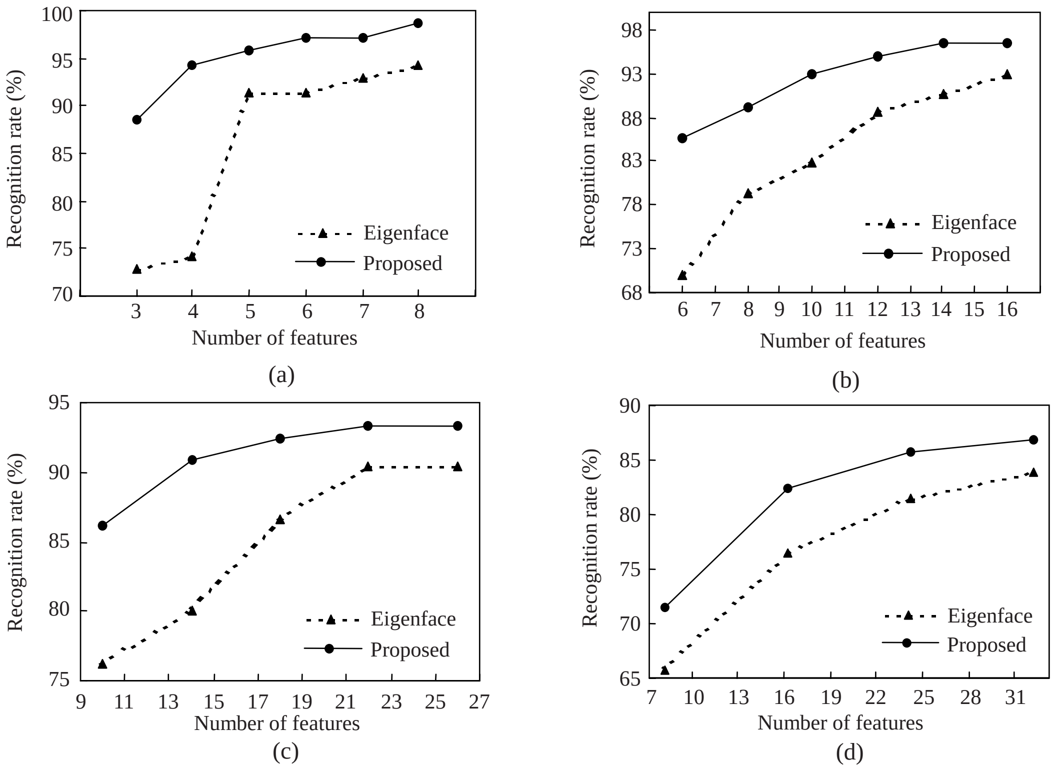


Fig.7 Recognition rate for different numbers of features in Case 1
(a) 10 classes; (b) 20 classes; (c) 30 classes; (d) 40 classes

recognition system rises and eventually becomes constant after a certain number is reached. Also, the proposed method generated much better results than the standard eigenface method did. Table 2 indicates that the recognition rate for 10 classes is 98.57% for 8 selected features in which one feature has been repeated. This shows the efficiency of the proposed method. The eigenface method would only yield a maximum recognition rate of 97.14%, even if we had used all the 30 features. As seen from Table 3, the advantage of the proposed method becomes evident. For instance, the proposed method fulfils a 100% recognition rate (10 classes) for a subset of features that would have never been possible even if we had used all the features in the eigenface method.

Table 3 shows another major advantage of the proposed method. Apparently, a much better recognition rate is achieved with a smaller number of features; e.g., for 40 classes, in the eigenface method, we have a recognition rate of 85.71% for 120 features, whereas using the proposed method we obtain a higher rate (86.79%) for only 32 features.

Experiment 2 This experiment was performed on 5 images per person as the training images and the remaining 5 as the test images for different numbers of classes.

Table 4 indicates the minimum numbers of features that yield 100% recognition. Table 5 shows the subsets of features selected that yield the maximum recognition rate for different numbers of classes. We see that the number of features as selected by the proposed method is much fewer than that required for the eigenface method to produce the same 100% result. From Figs.8a~8d and Table 4, it is clear that the proposed method completely outperforms the eigenface method in this experiment.

Experiment 3 Another program in MATLAB was written to evaluate, compare and contrast the response of PCA extended with GA (PCA-GA) with the proposed PCA-MA method for 40 classes (Cases 1 and 2). PCA-MA has been found to perform far better than the PCA-GA method (Figs.9a and 9b).

From Figs.9a and 9b one could safely infer that with the increase of features, PCA-GA approaches PCA-MA in performance. Under all other circumstances, PCA-MA is superior to PCA-GA.

Table 3 Number of features giving the maximum recognition rate (Case 1, ORL)

Method	Number of features (eigenvectors)				Maximum recognition rate (%)			
	n=10	20	30	40	*n*=10	20	30	40
Eigenface	30	60	90	120	97.14	94.29	90.95	85.71
Proposed	17	36	22	32	100.00	97.14	93.33	86.79

n: number of classes

Table 4 Number of features giving the maximum recognition rate (Case 2, ORL)

Method	Number of features (eigenvectors)				Maximum recognition rate (%)			
	n=10	20	30	40	*n*=10	20	30	40
Eigenface	13	20	51	180	100	100	100	100
Proposed	6	18	34	48	100	100	100	100

n: number of classes

Table 5 Subsets of features selected (Case 2, ORL)

Number of classes	Selected features (eigenvectors)	Number of selected features	Maximum recognition rate (%)
10	(3,4,5,6,7,8)	6	100
20	(1,2,3,4,5,7,9,10,12,13,14,15,17,18,19,21,22,25)	18	100
30	(1,2,3,4,5,7,8,9,10,11,12,13,14,15,16,17,18,20,20,21,22,23,25,26, 27,30,32,35,35,37,39,40,42,43)	34	100
40	(1,3,4,5,6,7,8,9,10,12,13,14,15,16,17,18,19,20,21,22,23,25,27,28, 29,31,33,34,35,36,37,38,39,40,41,42,45,46,47,48,49,50,51,55,56, 58,59,60)	48	100

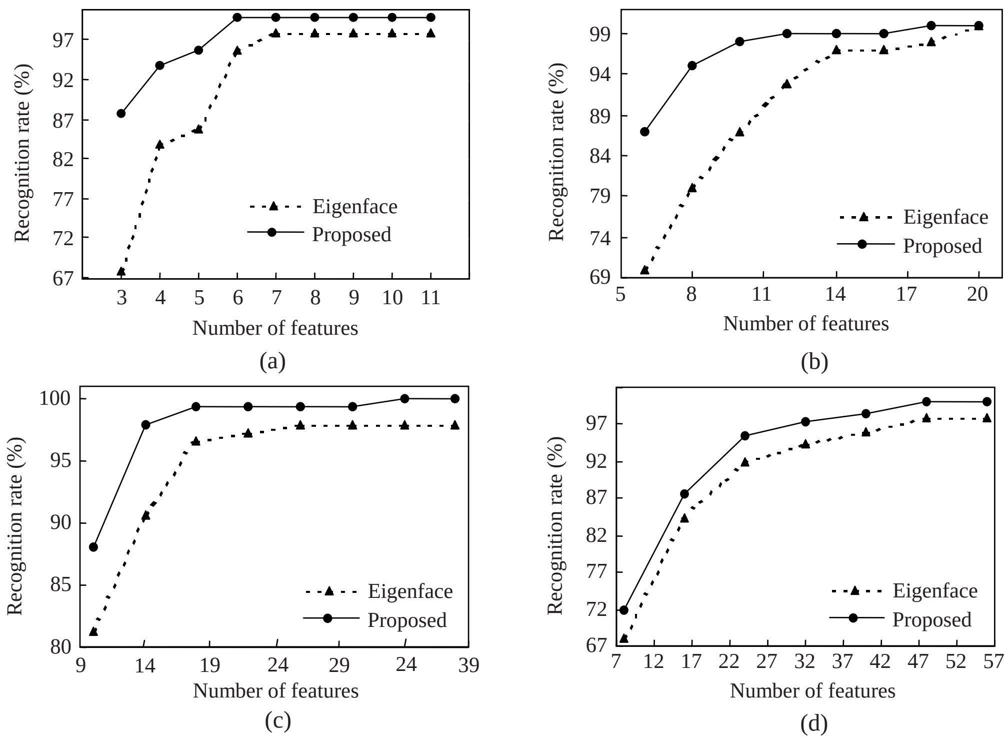


Fig.8 Recognition rate for different numbers of features in Case 2
(a) 10 classes; (b) 20 classes; (c) 30 classes; (d) 40 classes

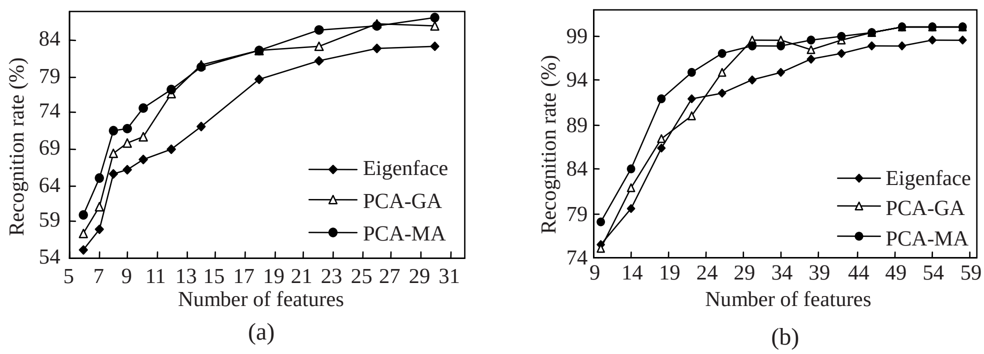


Fig.9 Recognition rate for different numbers of features. (a) Case 1; (b) Case 2

Experiment 4 The proposed method was further applied on the YaleB face database. Figs.10a~10d show the results. This experiment was also performed for both of Cases 1 and 2.

Fig.10 shows that the proposed method is evidently superior to the eigenface method.

Experiment 5 This experiment was performed to compare the results of the PCA-MA method with those of LDA-MA and KPCA-MA methods. These methods used the same memetic approach for feature (eigenvector) selection out of the eigenvectors as obtained by using LDA and KPCA methods. It is known that, in LDA the dimension is finally reduced to $C-1$, where C is the number of classes. For KPCA, the polynomial kernel $k(x_i, x_j)=(x_i, x_j)^d$ of degree d (d=2, 3, 4) has been used.

Fig.11 highlights the performance of various approaches in terms of the recognition rate vs the number of features. The performance of various approaches in terms of the number of features is tabulated in Table 6 with the best results highlighted. Table 7 summarizes the performance of various approaches in terms of the minimum number of features that produce a maximum recognition rate, and the corresponding subsets of features are given in Table 8.

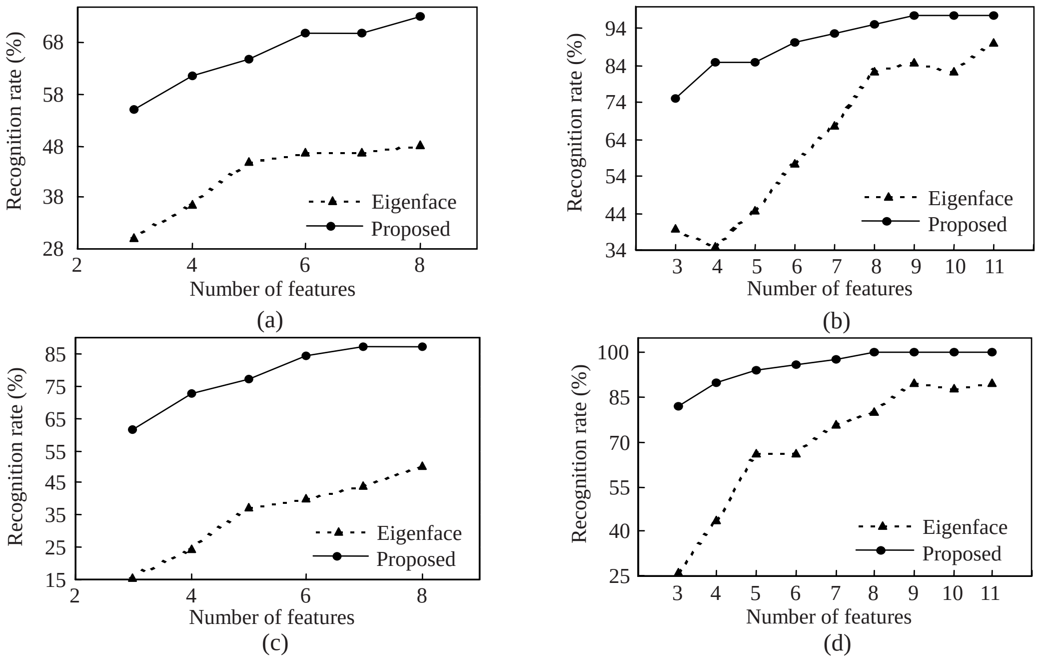


Fig.10 Recognition rate for different numbers of features

(a) YaleB_Pose, Case 1; (b) YaleB_Pose, Case 2; (c) YaleB_Illumination, Case 1; (d) YaleB_Illumination, Case 2

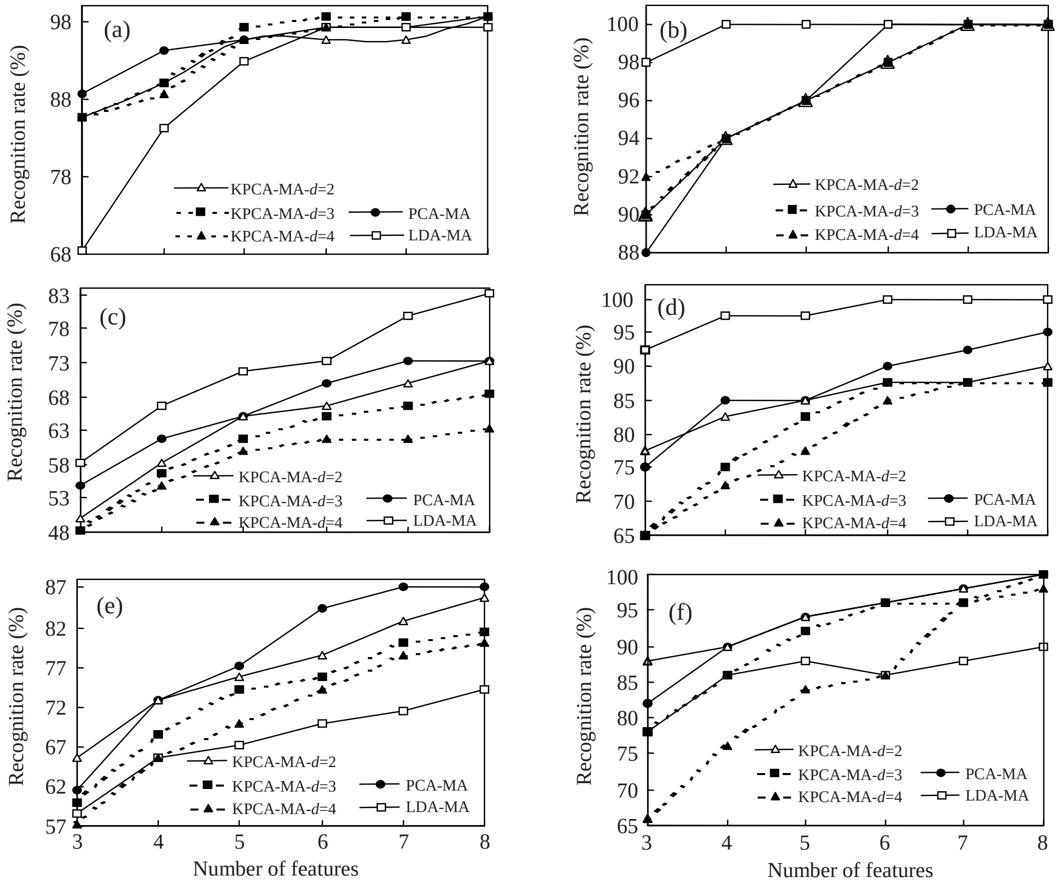


Fig.11 Recognition rate for different numbers of features

(a) ORL, Case 1; (b) ORL, Case 2; (c) YaleB_Pose, Case 1; (d) YaleB_Pose, Case 2; (e) YaleB_Illumination, Case 1; (f) YaleB__Illumination, Case 2

Table 6 Recognition rate for a fixed number of features

Method	Recognition rate (%)					
	Case 1			Case 2		
	ORL	YBP	YBI	ORL	YBP	YBI
Eigenface (PCA)	94.29	48.33	50.00	84.00	57.50	80.00
PCA-MA	**98.57**	73.33	**87.14**	94.00	90.00	**100.00**
LDA	87.14	73.33	60.00	70.00	87.50	84.00
LDA-MA	97.14	**83.33**	74.29	**100.00**	**100.00**	90.00
KPCA						
d=2	94.29	36.67	67.14	82.00	52.50	72.00
d=3	94.29	36.67	65.71	82.00	52.50	62.00
d=4	94.29	33.33	62.86	82.00	50.00	52.00
KPCA-MA						
d=2	**98.57**	73.33	85.71	94.00	87.50	**100.00**
d=3	**98.57**	68.33	81.42	94.00	87.50	**100.00**
d=4	**98.57**	63.33	80.00	94.00	85.00	98.00

The number of features is 8 for all the three databases in Case 1 and YaleB_Illumination in Case 2, 4 for ORL in Case 2, and 6 for YaleB_Pose in Case 2. YBP: YaleB_Pose, YBI: YaleB_Illumination. The bold numbers are the best results

Table 7 Minimum number of features with maximum recognition rate

Method	Minimum number of features						Maximum recognition rate (%)					
	Case 1			Case 2			Case 1			Case 2		
	ORL	YBP	YBI	ORL	YBP	YBI	ORL	YBP	YBI	ORL	YBP	YBI
Eigenface (PCA)	10	18	29	13	15	16	97.14	60.00	81.43	100.00	95.00	98.00
PCA-MA	17	9	**10**	6	14	**8**	100.00	83.33	**95.71**	100.00	100.00	**100.00**
LDA	8	8	9	9	8	8	87.14	73.33	70.00	92.00	100.00	84.00
LDA-MA	8	**8**	8	**4**	**6**	8	97.14	**83.33**	74.29	**100.00**	**100.00**	90.00
KPCA												
d=2	26	24	27	21	35	23	97.14	55.00	78.57	100.00	82.50	100.00
d=3	26	22	22	23	36	22	97.14	53.33	78.57	100.00	82.50	100.00
d=4	26	17	22	32	42	27	97.14	53.33	78.57	100.00	82.50	100.00
KPCA-MA												
d=2	**9**	28	20	7	10	**8**	**100.00**	81.67	94.29	100.00	95.00	**100.00**
d=3	10	25	20	7	12	**8**	100.00	78.00	91.43	100.00	92.50	**100.00**
d=4	11	22	28	7	20	10	100.00	80.00	90.00	100.00	92.50	100.00

YBP: YaleB_Pose, YBI: YaleB_Illumination. The bold numbers are the best results

Table 8 Subset of features selected (Cases 1 and 2)

Method and database	Selected features (eigenvectors)	Number of selected features	Maximum recognition rate (%)
KPCA+MA (*d*=2) (ORL, Case 1)	(5,14,2,13,8,9,6,1,4)	9	100.00
LDA+MA (YaleB_Pose, Case 1)	(8,7,5,8,1,4,6,8)	8	83.33
PCA+MA (YaleB_Illumination, Case 1)	(3,16,6,9,5,11,8,10,7,4)	10	95.71
LDA+MA (ORL, Case 2)	(7,1,2,9)	4	100.00
LDA+MA (YaleB_Pose, Case 2)	(6,9,5,1,2,4)	6	100.00
PCA+MA (YaleB_Illumination, Case 2)	(12,5,3,4,11,6,13,9)	8	100.00
KPCA+MA (*d*=2) (YaleB_Illumination, Case 2)	(10,4,28,3,6,18,16,14)	8	100.00
KPCA+MA (*d*=3) (YaleB_Illumination, Case 2)	(9,10,4,8,14,19,17,29)	8	100.00

DISCUSSION AND CONCLUSION

This paper proposes a PCA-MA method of feature selection for face recognition via the extension of PCA by adding MA. We evaluated the response of the PCA-MA method through experiments, and observed that the recognition rate based on features selected by the proposed method is much better than that obtained by the eigenface method for the same number of features. The recognition rate for different numbers of selected features (eigenvectors) such as 3, 4, 5, 6, 7, 8 for 10 classes (Case 1) was better than that in the eigenface method. The performance of the Eigenface method improved as the number of features increased; but still it was found to be superior to that of the PCA-MA method. This is true for all classes in both of the cases. Furthermore, as shown in Table 3, the eigenface method could produce a 97.14% recognition rate with 30 features whereas the PCA-MA method produced a 100% recognition rate with only 17 features. The 100% rate could not be obtained even with all the features used in the eigenface method. In addition, Table 3 clearly indicates that the PCA-MA method is better on all counts of observations. As compared with the PCA-GA method, our PCA-MA method also yields better results (Fig.9).

As the performance comparison of six methods (Table 6) reveals, in Case 1, for a fixed number of features, PCA-MA and KPCA-MA produced the best results for the ORL database, whereas for the YaleB face database, LDA-MA for YaleB_Pose and PCA-MA for YaleB_Illumination gave the best performance. In Case 2, LDA-MA for ORL and YaleB_Pose databases gave the best results, whereas for YaleB_Illumination PCA-MA and KPCA-MA (d=2, 3) achieved equal performance. From Table 7, it is clear that KPCA-MA (d=2) for ORL, LDA-MA for YaleB_Pose and PCA-MA for YaleB_Illumination obtained the best results for Case 1. In Case 2, LDA-MA yielded the best results for ORL and YaleB_Pose, whereas for YaleB_Illumination the best performance was achieved by PCA-MA and KPCA-MA (d=2, 3). Our experimental results revealed that the addition of MA to PCA, LDA and KPCA contributes to a more efficient selection of features and stronger capacity of producing good results in terms of the recognition rate as compared to PCA, LDA and KPCA, respectively.

References

Aydemir, M.E., Gunel, T., Kargin, S., Erer, I., Kurnaz, S., 2005. SAR Image Processing by a Memetic Algorithm. Proc. 2nd Int. Conf. on Recent Advances in Space Technologies, p.684-687. [doi:10.1109/RAST.2005.1512654]

Bartlett, M.S., 2001. Face Image Analysis by Unsupervised Learning. Kluwer Academic Publishers, Boston.

Bartlett, M.S., Lades, H.M., Sejnowski, T.J., 1998. Independent Component Representations for Face Recognition. SPIE, **3299**:528-539. [doi:10.1117/12.320144]

Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J., 1997. Eigenface vs Fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.*, **19**(7):711-720. [doi:10.1109/34.598228]

Bonabeau, E., Dorigo, M., Theraulaz, G., 1999. Swarm Intelligence: From Natural to Artificial Systems. Oxford University Press, New York.

Chellappa, R., Wilson, C., Sirohey, S., 1995. Human and machine recognition of faces: a survey. *Proc. IEEE*, **83**(5):705-740. [doi:10.1109/5.381842]

Dash, M., Liu, H., 1997. Feature selection for classification. *Int. J. Intell. Data Anal.*, **1**:131-156. [doi:10.1016/S1088-467X(97)00008-5]

Dawkins, R., 1976. The Selfish Gene. Oxford University Press, New York.

Dorigo, M., Caro, G.D., 1999. Ant Colony Optimization: A New Meta-heuristic. Proc. Congress on Evolutionary Computation, **2**:6-9. [doi:10.1109/cec.1999.782657]

Dorigo, M., Stutzle, T., 2004. Ant Colony Optimization. MIT Press, Cambridge, USA.

Elbeltagi, E., Hegazy, T., Grierson, D., 2005. Comparison among five evolutionary-based optimization algorithms. *Adv. Eng. Inf.*, **19**(1):43-53. [doi:10.1016/j.aei.2005.01.004]

Georghiades, A., Belhumeur, P., Kriegman, D., 2001. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.*, **23**(6):643-660. [doi:10.1109/34.927464]

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Publishing Co., MA.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. *Mach. Learn.*, **46**(1-3):389-422. [doi:10.1023/A:1012487302797]

Holz, H.J., Loew, M.H., 1994. Relative Feature Importance: A Classifier Independent Approach to Feature Selection. *In*: Gelsema, E.S., Kanal, N.L. (Eds.), Pattern Recognition in Practice IV. Amsterdam, Elsevier, p.473-487.

Hotelling, H., 1993. Analysis of a complex of statistical variables into principal components. *J. Educat. Psychol.*, **24**:417-441, 498-520.

Hsu, C.N., Huang, H.J., Dietrich, S., 2004. The ANNIGMA-wrapper approach to fast feature selection for neural nets. *IEEE Trans. Syst., Man, Cybern. B*, **32**(2):207-212. [doi:10.1109/3477.990877]

Jain, A., Zongker, D., 1997. Feature selection: evaluation, application, and small sample performance. *IEEE Trans. Pattern Anal. Mach. Intell.*, **19**(2):153-158. [doi:10.1109/34.574797]

Kim, K.I., Jung, K., Kim, H.J., 2002. Face recognition using kernel principal component analysis. *IEEE Signal Process. Lett.*, **9**(2):40-42. [doi:10.1109/97.991133]

Kirby, M., Sirovich, L., 1990. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. Pattern Anal. Mach. Intell.*, **12**(1):103-108. [doi:10.1109/34.41390]

Kittler, J., 1978. Feature Set Search Algorithms. *In*: Chen, C.H. (Ed.), Pattern Recognition and Signal Processing. Sijthoff and Noordhoff, The Netherlands, p.41-60.

Kohavi, R., John, G.H., 1997. Wrapper for feature subset selection. *Artif. Intell.*, **97**(1/2):273-324. [doi:10.1016/S0004-3702(97)00043-X]

Kwak, N., Choi, C.H., 2002. Input feature selection by mutual information based on parzen window. *IEEE Trans. Pattern Anal. Mach. Intell.*, **24**(12):1667-1671. [doi:10.1109/TPAMI.2002.1114861]

Lee, K.C., Ho, J., Kreigman, D., 2005. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intell.*, **27**(5):684-698. [doi:10.1109/tpami.2005.92]

Mao, K.Z., 2004. Feature subset selection for support vector machines through discriminative function pruning analysis. *IEEE Trans. Syst., Man, Cybern. B*, **34**(1):60-67. [doi:10.1109/TSMCB.2002.805808]

Merz, P., Freisleben, B., 1997. A Genetic Local Search Approach to the Quadratic Assignment Problem. Proc. 7th Int. Conf. on Genetic Algorithms, p.465-472.

Merz, P., Freisleben, B., 2000. Fitness landscape analysis and the memetic algorithms for the quadratic assignment problem. *IEEE Trans. Evol. Comput.*, **4**(4):337-352. [doi:10.1109/4235.887234]

Moscato, P., 1999. Memetic algorithms: A Short Introduction. *In*: Corne, D., Dorigo, M., Glover, F. (Eds.), New Ideas in Optimization. McGraw-Hill, Maidenhead, UK, p.219-234.

Narendra, P.M., Fukunaga, K., 1977. A branch and bound algorithm for feature subset selection. *IEEE Trans. Comput.*, **C-26**:917-922. [doi:10.1109/TC.1977.1674939]

Oh, I.S., Lee, J.S., Moon, B.R., 2004. Hybrid genetic algorithms for feature selection. *IEEE Trans. Pattern Anal. Mach. Intell.*, **26**(11):1424-1437. [doi:10.1109/TPAMI.2004.105]

Pudil, P., Novovicova, J., Kittler, J., 1994. Floating search methods in feature selection. *Pattern Recogn. Lett.*, **15**(11):1119-1125. [doi:10.1016/0167-8655(94)90127-9]

Schölkolf, B., Smola, A., Müller, K.R., 1998. Nonlinear component analysis as a kernel eigenvalue problem. *Neur. Comput.*, **10**(5):1299-1319. [doi:10.1162/089976698300017467]

Sheng, W.G., Howells, G., Fairhurst, M., Deravi, F., 2007. A memetic fingerprint matching algorithm. *IEEE Trans. Inf. Forens. Secur.*, **2**(3):402-412. [doi:10.1109/TIFS.2007.902681]

Sheng, W.G., Liu, X.H., Fairhurst, M., 2008. A niching memetic algorithm for simultaneous clustering and feature selection. *IEEE Trans. Knowl. Data Eng.*, **20**(7):868-879. [doi:10.1109/TKDE.2008.33]

Siedlecki, W., Sklansky, J., 1989. A note on genetic algorithms for large scale feature selection. *Pattern Recogn. Lett.*, **10**(5):335-347. [doi:10.1016/0167-8655(89)90037-8]

Swets, D.L., Weng, J.J., 1996. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.*, **18**(8):831-836. [doi:10.1109/34.531802]

Turk, M., Pentland, A., 1991. Eigenfaces for recognition. *J. Cogn. Neurosci.*, **3**(1):71-86. [doi:10.1162/jocn.1991.3.1.71]

Xing, E., Jordan, M., Karp, R., 2001. Feature Selection for High-dimensional Genomic Microarray Data. Proc. 15th Int. Conf. on Machine Learning, p.601-608.

Yan, Z., Yuan, C., 2004. Ant colony optimization for feature selection in face recognition. *LNCS*, **3072**:221-226.

Yang, M.H., Ahuja, N., Kreigman, D., 2000. Face Recognition Using Kernel Eigenfaces. Proc. Int. Conf. on Image Processing, **1**:37-40. [doi:10.1109/icip.2000.900886]

Zhao, W., Chellapa, R., Rosenfeld, A., Phillips, P.J., 2003. Face recognition: a literature survey. *ACM Comput. Surv.*, **35**(4):399-458. [doi:10.1145/954339.954342]

Zhu, Z., Ong, Y.S., 2007. Memetic algorithms for feature selection on microarray data. *LNCS*, **4491**:1327-1335. [doi:10.1007/978-3-540-72383-7_155]

Zhu, Z., Ong, Y.S., Dash, M., 2007. Wrapper–filter feature selection algorithm using a memetic framework. *IEEE Trans. Syst., Man, Cybern. B*, **37**(1):70-76. [doi:10.1109/TSMCB.2006.883267]