Digital Signal Processing **12**, 21–46 (2002)
doi:10.1006/dspr.2001.0413, available online at http://www.idealibrary.com on IDEAL®

An Adaptive Combined Classifier System for Invariant Face Recognition

G. A. Khuwaja

Department of Physics, Kuwait University, P.O. Box 5969, Safat 13060, Kuwait

Khuwaja, G. A., An Adaptive Combined Classifier System for Invariant Face Recognition, *Digital Signal Processing* **12** (2002) 21–46.

In classification tasks it may be wise to combine observations from different sources. In this paper, to obtain classification systems with both good generalization performance and efficiency in space and time, a learning vector quantization learning method based on combinations of weak classifiers is proposed. The weak classifiers are generated using automatic elimination of redundant hidden layer neurons of the network on both the entire face images and the extracted features: forehead, right eye, left eye, nose, mouth, and chin. The neuron elimination is based on the killing of blind neurons, which are redundant. The classifiers are then combined through majority voting on the decisions available from input classifiers. It is demonstrated that the proposed system is capable of achieving better classification results with both good generalization performance and a fast training time on a variety of test problems using a large and variable database. The selection of stable and representative sets of features that efficiently discriminate between faces in a huge database is discussed.

Key Words: pattern recognition; face recognition; adaptive classification; learning vector quantization; combined classifier; invariant recognition; feature selection.

1. INTRODUCTION

Pattern recognition is concerned with the automatic detection and classification of objects or events. Emerging new applications resulted in increasing interest in pattern recognition. The interest in pattern recognition is so diverse that pattern recognition applications attract researchers belonging to all possible disciplines ranging from business to the most sophisticated industries. Invariant face classification is a challenging task, especially in the absence of highly controlled environments and recognition constraints. Recent progress of computer technology has made us expect the face will play a key role in future

1051-2004/02 $35.00

human–machine interaction and advanced communications, such as multimedia and low-bandwidth video-telephony.

Two of the most important issues for supervised learning can be specified as generalization performance and efficiency. The former addresses the problem of how to develop an adaptive pattern recognition system to achieve optimal performance on samples that are not included in a training set with a finite number of training samples. The latter deals with the complexity of a pattern recognition system in both space and time. The space complexity refers to the size of a system and the time complexity characterizes the computational time needed to develop such systems.

Tremendous efforts have been made on estimating and finding the optimal architecture of a classification system using a finite number of training samples. These approaches include computational learning theory in both the machine learning and neural-network community [1–6] and various statistical methods including cross-validation and model selection [7–9]. Since it is very difficult to find the best architecture [10], methods are proposed to combine different architectures. Specifically, in the pattern recognition community, combinations of classifiers are proposed to improve the classification performance of a single classifier [11–13]. Also, it had been observed in design studies that although one of the designs would yield the best performance, the sets of patterns misclassified by different classifiers would not necessarily overlap. This suggested that different classifier designs potentially offered complementary information about the patterns to be classified, which could be harnessed to improve the performance of the selected classifier.

These observations motivated the relatively recent interest in combining classifiers. The idea is not to rely on a single decision making scheme. Instead, all the designs or their subset are used for decision making by combining their individual opinions to derive a consensus decision. Various classifier combination schemes have been devised and it has been experimentally demonstrated that some of them consistently outperform a single best classifier.

Commonly, a combined decision is obtained by just averaging the estimated posterior probabilities. This simple algorithm already gives very good results [14, 15]. This result is somewhat surprising especially considering the fact that averaging of the posterior probabilities is not based on some solid (Bayesian) foundation [16].

A large number of combining schemes for classification exist. In general three types of situations in combining classifiers can be identified [17]. In the first type each classifier outputs a single class label and these labels have to be combined [18]. In the second type the classifiers' output sets of class labels ranked in the order of likelihood [19] and the third type involves the combination of real valued outputs for each class by the respective classifiers [20, 21]. This research work incorporates the first two combination schemes.

The problem of combination of classifiers is handled by generating a number of weak classifiers based on automatic elimination of redundant hidden layer neuron networks on both the entire face images and the extracted features: forehead, right eye, left eye, nose, mouth, and chin. The neuron elimination is based on the pruning of blind neurons, which are redundant. Redundancy of the neuron is measured by the variance of the face image represented by that neuron. The classifiers are then combined through majority voting and ranking level on the decisions available from input classifiers. Section 2 gives a brief overview of previous face recognition techniques. The proposed system is tested on the AR Database [22]. Section 3 describes the learning vector quantization (LVQ) classifier. Section 4 presents the AR database used for this research work and the preprocessing of face images. The merits and demerits of combination schemes for the selection of the base model of combined classifiers are discussed in Section 5. The proposed learning algorithm for the generation of efficient classifiers based on LVQ models is discussed in Section 6. The empirical results are presented and discussed in Section 7. Finally, in Section 8, the paper is concluded.

2. OVERVIEW

Face recognition approaches could be categorized into two major categories: holistic approaches and feature-based approaches. Figure 1 shows the different subcategories of each approach. A detailed overview of face recognition approaches can be found in extensive surveys [23–25].

2.1. *Holistic Approaches for Face Recognition*

The eigenface approach described by Turk and Pentland [26] is one of the most popular approaches for face recognition. The principal component analysis is applied on the training set of faces. The eigenface approach assumes that the set of all possible face images occupies a low-dimensional subspace, derived from the original high-dimensional input image space. The eigenface space is an approximation of face patterns in the training set using data clusters and

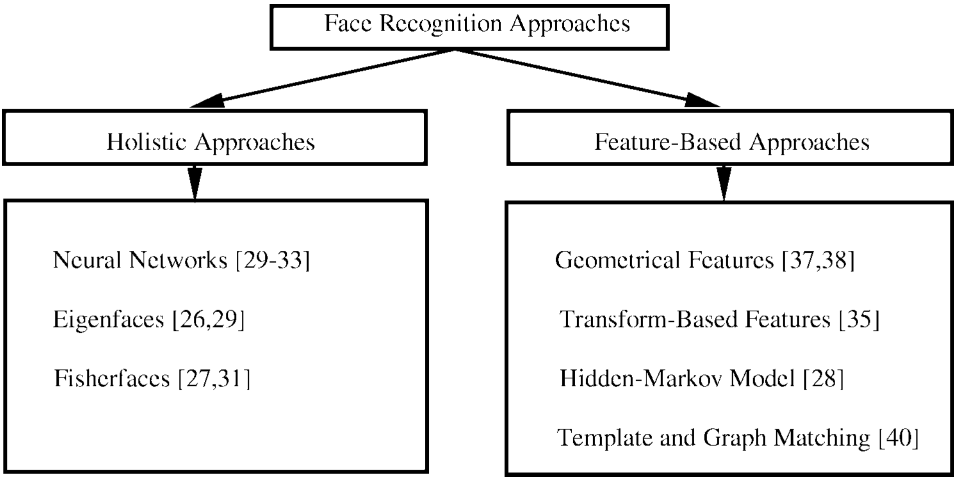

FIG. 1. Face recognition techniques.

their principal components. An unknown face is classified if its distance to the clusters is below a certain threshold, using an appropriate classifier. Turk and Pentland [26] reported a correct recognition rate of 95% in the case of the FERET database containing about 3000 different faces. The tested face images seem to be taken with little variation in viewpoint and lighting, although with significant variation in facial expression.

The major drawback of the eigenface approach [27] is that the scatter being maximized is due not only to the between-class scatter that is useful for classification, but also to the within-class scatter that, for classification purposes, is unwanted information. Much of the variation from one image to the next is due to illumination changes. Thus if PCA (principal component analysis) is presented with images of faces under varying illumination, the projection matrix W_{opt} will contain principal components (i.e., eigenfaces) which retain, in the projected feature space, the variation due to lighting. Consequently, the points in the projected space will not be well clustered, and worse, the classes may be smeared together. It has been suggested [27] that by discarding the three most significant principal componenets, the variation due to lighting is reduced.

Many other researchers have implemented the eigenface approach for comparison purposes. Belhumeur *et al.* [27] used the Fisherface method to project face images into a three-dimensional linear subspace. The projection is based on Fisher's linear discriminant in order to maximize the between-class scatter while minimizing the within-class scatter. This approach is proved to be more efficient than the eigenface approach specially in the case of variable illumination. The experiments were performed on only 150 faces from 15 subjects selected from the ORL database [28]. The results show that the eigenface approach is quite robust when dealing with glasses and facial expressions but sensitive to scale, pose, and illumination. The correct recognition rate achieved is 95% for only 150 images selected from the 400 images of the ORL database.

In Lawrence *et al.* [29], testing the eigenface method on the ORL database resulted in 89.5% correct recognition rate. Both a convolutional neural network and a self-organizing feature map classifier were used for invariant face recognition. This system was tested on the ORL database and resulted in a correct recognition rate of 96.2% for the case of a training set including five faces per person and a test set including five faces per person.

In Lin *et al.* [30], a probabilistic decision-based neural network (PDBNN) is described for face recognition. While the system performance in the case of the FERET database is 99%, its performance for the case of the ORL database is 94%. The face recognition time is less than 0.1 s on an SGI Indy machine.

In Feitosa *et al.* [31], the performance of both the linear discriminant analysis (LDA) and a Gaussian mixture model that is based on the radial basis function (RBF) network is compared. Experiments are performed on the ORL database. The database is divided into a training set and a testing set. Each set includes 200 randomly selected images (5 images $\times$ 40 subjects). For implementation convenience all images were first resized to 64×64 pixels. Each image is then represented by one vector, which is obtained by simply concatenating the columns of the image matrix. PCA is first used for dimensionality reduction by keeping only the most significant 50 eigenfaces. Both LDA and RBF classifiers are then trained on the most significant eigenfaces. The results indicate that the more general model underlying the RBF classifier does not bring any significant performance compared with the LDA approach. The best average recognition rate (95.5%) of the RBF approach was obtained for 50 eigenfaces working with 110 neurons in the RBF hidden layer. The average recognition rate for the LDA was 95.7% when using 39 most discriminant features (MDFs) computed on 50 most expressive features (MEFs). The ability of the RBF network to use more than one Gaussian to describe the population of each group brought no significant performance improvement when compared to the less computation intensive LDA classifier. Training the classifiers on facial feature vectors does not consider the textural characteristics of the face. Using the RBF classifier with only 110 hidden neurons means less generalization for variability. More neurons mean a better chance to encompass a wider range of poses and scales. In addition to that, the training set is selected randomly which could lead to unlearned variability situations.

In Srinivas and Wechsler [32], a hybrid architecture is used for forensic classification and retrieval tasks. The classifier consists of an ensemble of RBF networks and inductive decision trees. Experimental results proving the feasibility of the approach yielded 96% accuracy for surveillance, using a database consisting of 904 images corresponding to 350 subjects. It has been shown that when the connectionist ensemble RBF (ERBF) model is coupled with the inductive decision tree, the performance improves over the case when only the ERBF module is used.

In Zhang and Flucher [33], a tree of neural networks is described for translation invariant face recognition. The problem of classifier fusion is addressed. The tree is capable of handling large databases with a large number of classes and noisy inputs and is capable of being upgraded to recognize new tasks without the need for retraining. The face recognition system locates the captured faces using MLP neural networks. A GAT (group-based adaptive tolerance) tree is used in the middle level for face recognition, using normalized face images. The face is classified as a front face, tilted to the left, tilted to the right, rotated to the left, or rotated to the right. Successful classification is followed by recognition of translation invariant faces by adaptively growing new nodes in the GAT tree in tolerance space. Simultaneously, faces with glasses and/or beards are classified using the same (GAT tree) technique. A higher level of recognition used neural networks, fact bases, rule bases, knowledge bases, and reasoning networks to perform more intelligent recognition. Each node in the tree consists of a neural-network group. Experiments were performed on 28×28 gray level images. For front face recognition, 87 different perspective faces were chosen as training exemplars and 693 faces for testing purposes. The GAT tree tests resulted in only one error case, which corresponds to an error rate of only 0.15%. Similar experiments were carried out for tilted and rotated face recognition. After training, GAT trees were able to recognize tilted and rotated faces with similar confidence levels. Training is performed on 136 faces and tested with 653 faces. The observed errors for tilted and rotated face recognition were 0.16 and 0.31%, respectively. To recognize front beard faces, six faces were chosen to train each GAT node (three front beard faces, three nonbeard faces) with 70 faces reserved for testing. The outputs of the network group were all greater than 0.92 for the four front beard face test cases. For the remaining 66 people (not front beard faces), the outputs of one GAT tree node were less than 0.92. The tolerance of the GAT to noise was tested and is found to be more than that of the general trees.

In Chen *et al.* [34], a minimum classification error rate based face recognition system is proposed. The minimum classification error formulation is incorporated into a neural network classifier called a multilayer perceptron. Experimental results show that the proposed system is robust to noisy images and complex background.

2.2. Feature-Based Approaches for Face Recognition

In Samaria [28], a hidden Markov model (HMM) based method is described for the extraction of facial features. An image is first converted into a one-dimensional vector of pixel intensities. The intensity vector can be used to train the HMM, which will then partition the sequence into a number of feature states. An HMM is primarily characterized by a transition probability matrix A that records the transitions from one feature state to another and an output probability matrix B that records the probability of going from a state to itself. A trained HMM on a sequence captures different aspects of the face image. Both matrices provide strong discrimination for the various subjects. A separate HMM was trained for each subject in the ORL database and the resulting models were used to classify unknown images. The statistical features obtained by the HMM have been shown to correspond to physical features as understood by humans when structural information is used to build the model. Using five training faces per person resulted in 95% correct recognition.

In Hagen [35], a Fourier spectrum analysis technique is used for face recognition. Recognition is done by finding the closest match between feature vectors containing the Fourier coefficients at selected frequencies. A template based approach uses 27 Fourier coefficients to yield 98% correct recognition. The coefficients which encompass the highest variance are selected. Classification of the transform coefficients is performed using the Euclidean minimum distance classifier. Experiments were performed on the ORL face database. The author compared the proposed Fourier spectrum technique with the Euclidean classifier with both the backpropagation (BKP) neural network and the eigenface method. The three different techniques described previously indicate that the Fourier transform based system shows superior performance (98%) when compared to both the eigenface (94%) and BKP neural network (96.5%) methods.

In Ben-Arie and Nandy [36], a volumetric–iconic frequency domain representation (VFR) model based system is tested on the ORL database and resulted in 92.5% correct classification for the case of five training faces per individual.

Using eight training faces resulted in 100% correct recognition. The major drawback of this system is that a face can be recognized in 320 s.

In Kin-Man and Hong [37], an analytic-to-holistic approach is introduced for identification of faces at different perspective variations. The ORL database is used in the experiments. Only one upright frontal face is selected for each of 40 individuals. Among the rest of the faces, they selected 160 images as a testing set. About half of the faces are upright and have a small rotation on the y-axis. The other half show different amounts of perspective variations. Fifteen feature points are located on a face. A head model is proposed and the rotation of the face can be estimated using geometrical measurements. The positions of the feature points are adjusted so that their corresponding positions for the frontal view are approximated. A similarity transform is then used to compare the feature points with prestored features. In addition, eyes, nose, and mouth are correlated with corresponding patterns in a database. Under different perspective variations, the overall recognition rates are over 84 and 96% for the first and the first three likely matched faces, respectively.

In Li and Lu [38] a classification method called the nearest feature line (NFL) is proposed for face recognition. The line passing through two feature points in the eigenspace of the same class is used to generalize any two feature points of the same class. The derived FL can capture more variations of face images than the original points. A nearest distance based classifier is used. The nearest feature line method achieved an error rate of 3.125%. The authors expect this improvement to be due to the feature lines' ability to expand the representational capacity of available feature points and to account for new conditions not represented by original prototype face images. The error rate of the proposed method is 43.7–65.4% of that of the standard eigenface method.

In Sutherland *et al.* [39], an input image containing the entire face is broken up into eight features of interest: the eyes, the bridge of the nose, nostrils, mouth, chin, hair, and the entire face. The vector quantization (VQ) of the facial features is performed after these features have been extracted from the entire image. One vector quantizer is dedicated to each of the eight features used. The vector quantizers are used here for data reduction. The VQ process thus yields a set of indices for all eight features representing the most likely vectors used to code the subject face. The VQ algorithm was first trained on 300 images acquired from 30 subjects. Another 300 images were used for testing. The images were frontal face information only and the size and orientation were kept approximately constant throughout the experiment. Ten images of each person were used to construct the database of signatures. Some facial parts (entire face, hair, chin, and some parts of the nose) are spatially subsampled due to their relative unimportance in frontal face recognition. An additional image of each person was manually segmented and used to form the vector quantizer codebook for each feature. The facial features of a test vector are used to obtain a probability measure for those features belonging to a particular individual. A multiplicative accumulation is used to obtain the probability that all eight features present are a plausible representation of the individual under test. The highest probability score is then used to locate the most likely match for the test face. The data regarding facial interrelationships have not been integrated in the VQ coefficient analysis. The test results for 300 test images were 89.19%.

In Wiskott *et al.* [40], an elastic graph-matching algorithm is used with a neural network for face recognition. Faces are stored as flexible graphs or grids with characteristic visual features (Gabor features) attached to the nodes of the graph (labeled graphs). The Gabor features are based on the wavelet transform and have been shown to provide a robust information coding for object recognition (invariance against intensity or contrast changes). Furthermore, Gabor features are less affected by pose, size, and facial expression than raw gray level features. For image matching against a stored graph, the graph location in the image is optimized. It has been shown that elastic graph matching can successfully recognize faces from facial line drawings. The efficiency of the Gabor wavelets in recognizing line drawings is due to the fact that line drawings have dominant orientations of bars and step edges, and the Gabor code is also dominated by orientation features. Gender classification experiments performed on line drawings resulted in a correct-decision rate of better than 90%.

In Lin and Wu [41], an automatic facial feature extraction algorithm is presented. The algorithm is composed of two main stages: the face region estimation stage and the feature extraction stage. In the face region estimation stage, a second-chance region growing method is adopted to estimate the face region of a target image. In the feature extraction stage, genetic search algorithms are applied to extract the facial feature points within the face region. It is shown by simulation results that the proposed algorithm can automatically and exactly extract facial features with limited computational complexity.

In Yoon *et al.* [42], wavelet transform of the input 256×256 color image is performed and the input image is decomposed into low-pass and high-pass components. After finding the position of a face using the histogram of edges, a face region in low-pass band image is extracted. Since a RGB color image is easily affected by illumination, the image of the low-pass component is normalized, and a face region is detected using face color information. In this paper, 3000 images of 10 persons are used and KL transform is applied in order to classify face vectors effectively. The FCM (Fuzzy C-means) algorithm classifies face vectors which have similar features into the same cluster. In this case, the number of clusters is equal to that of a person, and the mean vector of each cluster is used as codebook. The recognition rate of learning images and that of testing images are computed using correlation coefficients and Euclidean distance.

3. LEARNING VECTOR QUANTIZATION CLASSIFIER

Learning vector quantization is a supervised classifier that was first studied by Kohonen [43]. Kohonen has proposed several variations on the basic LVQ algorithm. The most common are LVQ1, LVQ2, and LVQ3. All create decision regions that are near optimal. The basic LVQ classifier (LVQ1) divides the input space into disjoint regions. The decision boundaries created by LVQ1 have been

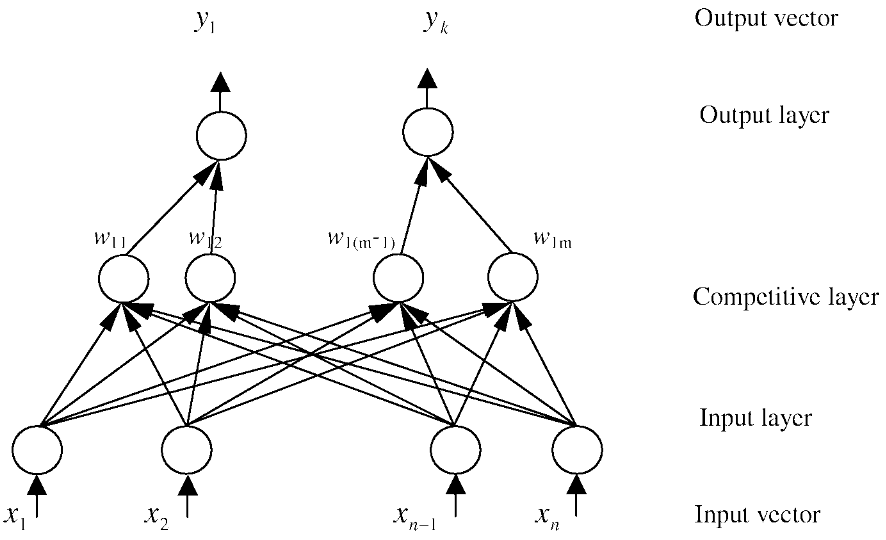

FIG. 2. Architecture of the LVQ classifier.

demonstrated to coincide closely with those of a Bayes classifier. A prototype vector represents each region. To classify an input vector, it must be compared with all prototypes. The Euclidean distance metric is used to select the closest vector to the input vector and the input vector is classified to the same class as the nearest prototype.

The LVQ classifier (Fig. 2) consists of an input layer, a hidden competitive layer, which learns to classify input vectors into subclasses, and an output layer, which transforms the competitive layer's classes into target classifications defined by the user. Only the winning neuron of the hidden layer has an output of one and other neurons have outputs of zero. The weight vectors of the hidden layer neurons are the prototypes, the number of which is usually fixed before training begins. The number of hidden neurons depends upon the complexity of the input–output relationship and significantly affects the results of classifier testing. Selection of the number of hidden neurons must be carefully made, as it highly depends on the encompassed variability in the input patterns. Extensive experiments are performed to conduct the suitable number.

For a training set containing n input faces, each of these faces is labeled as being one of k classes. The learning phase starts by initiating the weight vectors of neurons in the hidden layer. Then, the input vectors are presented randomly to the network. For each input vector X_j, a winner neuron W_i is chosen to adjust its weight vector:

$$\|X_j - W_i\| \leq \|X_j - W_k\|, \qquad \text{for all } k \neq i. \tag{1}$$

The weight vector $W_i(t)$ is updated to the next step $t + 1$ as follows:

$$W_i(t+1) = W_i(t) + \alpha(X_j - W_i(t)) \qquad \text{if } X_j \text{ and } W_i \text{ belong to the same class,} \tag{2}$$

$$W_i(t+1) = W_i(t) - \alpha(X_j - W_i(t)) \qquad \text{if } X_j \text{ and } W_i \text{ belong to different classes,} \tag{3}$$

where $0 \leq \alpha \leq 1$ is the learning rate, which may be kept constant during training or may be decreasing monotonically with time for better convergence [43]. Otherwise, do not change the weights. During the test phase, the distance of an input vector to each processing element of the hidden layer is computed and again the nearest element is declared the winner. This in turn fires one output neuron, signifying a particular class.

4. AR FACE DATABASE AND PREPROCESSING

Martinez and Benavente created the AR face database at the Computer Vision Center (CVC), Purdue University [22]. It contains over 4,000 color images corresponding to 136 people's faces (76 men and 60 women). This face database is publicly available and can be obtained from the Web site http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html. Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sunglasses and scarf). The pictures were taken under strictly controlled conditions. No restrictions on accessories (clothes, glasses, etc.), make-up, hairstyle, etc. were imposed on participants. Each person participated in two sessions, separated by two weeks (14 days) time. The same pictures were taken in both sessions. The various face features are (Fig. 3): neutral expression, smile, anger, scream, left light on, right light on, all side lights on, wearing sun glasses, wearing sun glasses and left light on, wearing sun glasses and right light on, wearing scarf, wearing scarf and left light on, wearing scarf and right light on. Figure 4 shows one facial image for 100 individuals (100 target classes) used in the training and testing for this study, i.e., a set of 500 facial images for training and another set of 500 facial images for testing, for a total of 1,000 facial images.

The investigations described in this paper are performed using entire facial images and the extracted features of the AR database: forehead, right eye, left eye, nose, mouth, and chin. The AR database images are stored as RGB RAW

FIG. 3. Various facial varieties of one person [22].

FIG. 4. Faces of 100 persons used for this study.

files (pixel information). Images are of 768×576 pixels and of 24 bits of depth. The whole set of images is resampled as 32×32, 48×48, and 64×64 and converted to 256 gray level images. To reduce the image size, a low pass filter is applied to the image before interpolation using the nearest (in a Euclidean distance sense) neighbor interpolation method. This reduces the effect of Moire patterns—ripple patterns that result from aliasing during resampling. After resampling all images have the same size.

5. COMBINED CLASSIFIERS

The interest in constructing combined classifiers for solving complex pattern recognition problems is increasing. Combining a set of classifiers can help solve the dilemma of bias and variance [44] and increase the efficiency and performance of the whole system. A combined model can be represented in a simple form. Let $h_k(x)$'s be a set of total K models defined on a feature vector $x \in R^d$, where $h_k(x)$ can be a neural network (or a single neuron). Let w_k be the weighting factor for the kth model. The combination of these K models can be represented [44] as

$$f_K(x) = \sum_{k=1}^{K} w_k h_k(x). \tag{4}$$

$f_K(x)$ is essentially a linear combination of the so-called base models $h_k(x)$'s. When a (two-class) classification is considered, the combined model becomes a combined classifier $C(x)$, where

$$C(x) = I(f_K(x)). \tag{5}$$

$I(z)$ is an indicator function, where $I(z) = 1$ for $z > 0$ and $I(z) = 0$ otherwise. $C(x)$ is a label assigned to a feature vector x by a combined classifier. When $h_k(x)$'s are also classifiers, $C(x)$ is a combination of classifiers. There are two key issues in combinations of models:

(1) How to obtain a set of base models, $\{h_k(x)\}_{k=1}^{K}$.

(2) Given a set of base models, how to choose an optimal set of weighting factors $\{w_k\}_{k=1}^{K}$ so that the generalization error of the combined model is minimized.

In [45], an analytical framework to quantify the improvements in classification results due to combining is provided. The results apply to both linear combiners and the order statistics combiners. Combining networks in output space reduces the variance of the actual decision region boundaries around the optimum boundary. For linear combiners, in the absence of classifier bias, the added classification error is proportional to the boundary variance. For nonlinear combiners, analytically the selection of the median, the maximum, and in general the ith order statistic improves classifier performance. Experimental results on several public domain data sets are provided to illustrate the benefits of combining.

It is well known that classifier combination provides increased accuracy over individual classifiers. Training multiple classifiers on different data sets can boost the performance of a single classification paradigm. Another combination scheme is to combine classifiers that use different features, training, and decision-making methodologies.

The ways of combining the outcomes from disparate classifiers are:

1. majority voting,
2. aggregation of ranked outputs (statistical theory of groups), and
3. front end supervised classifier.

Decision fusion is a recent trend in pattern recognition that aims at increasing the accuracy. Multiple decisions collected from the individual classifiers are integrated to improve the overall system accuracy. Fusion of multiple decisions may be done at three levels:

1. Abstract level: using voting techniques since the output of each classifier is only a set of possible labels without any confidence associated with the labels.

2. Rank level: the output of each module is a set of labels ranked by a decreasing confidence level.

3. Measurement level: the output from each module is a set of labels associated with confidence values.

In [46], recent advances in supervised learning are reviewed with a focus on the two most important issues: performance and efficiency. Performance addresses the generalization capability of a learning machine on randomly chosen samples that are not included in a training set. Efficiency deals with the complexity of a learning machine in both space and time. Four types of learning approaches are discussed: training an individual model; combinations of several well-trained models; combinations of many weak models; and evolutionary computation of models. The advantages and weaknesses of each approach and their interrelations are discovered.

The learning time must scale nicely with respect to the size of data sets. Since the size of learning machines determines the memory required for implementation, a learning machine with a compact structure is preferred. Therefore, a challenging problem is how to develop adaptive learning systems with a compact structure that can achieve good performance and be adapted in real time.

Several general frameworks, such as the expectation-maximization (EM) framework, the combination scheme, weak learning, and evolutionary algorithms, are discussed, all of which aim at improving the efficiency and performance of a learning machine. A neural network is used as a simple architecture to show how these general frameworks are applied. The reason why neural networks are chosen is that they have been shown to be universal approximators to a general class of nonlinear functions and have become popular recently.

The performance of a supervised learning system is characterized by its generalization error [44], which measures the distance between the output function of a trained model and an underlying target function. Most existing methods for training neural networks in supervised learning suffer from an intrinsic problem in pattern recognition: the bias and variance dilemma. That is, if a neural network is too large, it may over-fit a particular training set and thereby fails to maintain good generalization error. A small neural network, however, may be sufficient to approximate an optimal solution. In addition, one important algorithmic problem is how to deal with a complex optimization problem with possibly many local minima.

Combinations of weak classifiers and how combined weak classifiers make a tradeoff among performance, time complexity, and space complexity is discussed. The role of evolutionary computation in tackling the problems of performance and efficiency are introduced.

Both theoretical and empirical results [47] suggest that combinations of well-trained models can ease the bias–variance dilemma and improve the performance substantially. That is, we can select a base model with a relatively large size so that it has a small bias but a large variance. A combination scheme can be responsible for reducing the overall variance for a combined model.

In addition, as several models are combined, each of which may take a long time to train, an even longer training time results for a combination. Since training time is critical for real-time applications and is more difficult to tackle than the space complexity, a natural question to ask is whether combinations of models can be used to improve the time complexity at a reasonable cost of the space complexity. Combining weak models provides a promising answer to this question.

6. ALGORITHM

The ultimate goals in the design of a pattern recognition system are efficiency and performance. To achieve these goals, the design and implementation of a sophisticated hybrid pattern recognition system became a necessity. A hybrid system is a combination scheme of two or more subsystems. The most challenging step in the design of a pattern recognition system is the selection of a suitable base model which constitutes its building blocks. The next step is the features selection and extraction method. Careful selection of a feature extraction method highly simplifies the design of the classifier subsystem. The last step is the design of a suitable decision classifier; the output classifier subsystem that combines all the decisions available from the input classifiers using a particular methodology.

The algorithm consists of two steps: (1) generating individual weak classifiers through the elimination of redundant hidden layer neurons, and (2) combining a collection of weak classifier outputs available from both the same data patterns and the various extracted features through simple majority voting and ranking level.

6.1. Efficient LVQ Base Model

A generic learning vector quantization neural network consists of three layers. The first layer is the input layer, which consists of as many neurons as the number of input samples of the image to be recognized. The hidden layer size is problem dependent. The number of hidden layer neurons (NH) should be suitable to capture the knowledge of the problem domain. For example, training a neural network to recognize faces which belong to NC (number of classes) classes, at least NC hidden layer neurons are required. To capture a large range of input pattern variability, a large number of hidden layer neurons is necessary. But, the problem is how large should it be.

Normally, this number is overestimated by including excess neurons in the network. This means that after convergence, some neurons will be redundant in the sense that they do not evolve significantly and thus do not capture any data structure. Visualizing the learned pattern of the hidden layer neurons, it is found that there are neurons with completely blurred patterns, blind neurons, as these neurons did not see the faces which are clamped to the neurons of the input layer. Survival of the fittest neurons is then crucial to the progress of the learning process. Eliminating the blind neurons enhances the classifier performance. The algorithm for input classifiers based on an efficient LVQ model is as follows:

1. Initialize the network parameters:

Input layer size = Image size ($32 \times 32 = 1024$, $48 \times 48 = 2304$, or $64 \times 64 = 4096$ neurons).

Training set size = 20 or 50 or 100 persons × 5 faces = (100, 250, 500) facial images.

Number of classes NC = 20, or 50, or 100.

Number of hidden layer neurons = number of pattern classes (NC = 20, 50, or 100)* number of faces per person (5).

Learning rate = 0.7.

Set up the target vector which specifies the target class of each pattern in the training set.

Display update rate = 100.

Arrange the input patterns of the training set as one-dimensional columns in an array P.

Number of training epochs (EP) = 1000, 5000, 10,000.

Number of possible classifiers = 20, 50, or 100.

Number of blind neurons NB = 0.

2. Initialize an LVQ classifier:

   Initialization of the weight matrix for competitive layer w1.

   Initialization of the weight matrix for linear layer w2.

3. Start training of an LVQ classifier.

4. Test the trained classifier on both training and test sets and compute PCCTR and percentage of correct classification for the test set (PCCTS).

5. Detect the number of blind hidden layer neurons using the standard deviation (sdv) of the weight vector for each hidden layer neuron: If $\mathrm{sdv}(w_{1i}) < \{25, 45\}$ then increase NB.

6. Update the number of hidden layer neurons: $\mathrm{NH}(t+1) = \mathrm{NH}(t) - \mathrm{NB}(t)$. If $\mathrm{NH}(t+1) < \mathrm{NC}$ then $\mathrm{NH}(t+1) = \mathrm{NH}(t+1) + \mathrm{NC}$.

7. Check the stop criterion based on the required number of classifiers.

6.2. Threshold Level

The threshold level for the elimination of the redundant neurons is found experimentally using a trial and error method. For the case of entire facial images, the value of the standard deviation of the ith hidden layer neuron $\mathrm{sdv}(w_{1i})$ of 36 yields the best performance in terms of PCCTS. However, for the case of nose, it turned out to be $\mathrm{sdv}(w_{1i}) = 25$. This means that the threshold value varies depending on the nature of the object or feature region of the face.

6.3. Learning Rate

The convergence speed is directly related to the learning rate parameter (α). If α is small, the search path will closely approximate the gradient path, but convergence will be very slow due to the large number of update steps needed to reach a local minima. On the other hand, if α is large, convergence initially will be very fast, but the algorithm will eventually oscillate and thus not reach a minimum [48]. Various experiments are performed to get the value of α ranging from 0.005 to 1.0. The value of α that gives the best performace on the average of 10 runs of the multilayer feedforward network is 0.7 for entire facial images (Table 1).

TABLE 1

PCCTR of 10 Classifiers of Pruning Feed Forward Network for Various Values of α

CL	0.005	0.01	0.1	0.3	0.5	0.6	0.7	0.8	0.9	1.0
1	99	97	95	99	100	100	100	100	100	100
2	98	98	99	97	99	100	100	99	100	100
3	100	99	99	98	100	99	100	100	100	96
4	99	99	100	100	97	99	100	100	100	100
5	99	98	96	97	99	100	100	99	100	100
6	98	99	96	98	97	100	100	100	100	100
7	97	99	96	97	98	98	100	99	99	98
8	98	99	98	93	98	98	100	100	100	98
9	98	97	97	95	94	99	100	99	100	100
10	99	100	98	94	97	100	100	100	100	100

Note. PCCTR, percentage of correct classification of training set; α, learning rate; CL, classifier.

6.4. Feature Extraction

Humans can effortlessly recognize a familiar object under novel viewing conditions. This ability to generalize and deal efficiently with novel stimuli has long been considered a challenging example of brain-like computation that proved extremely difficult to replicate in artificial systems.

Features which represent a pattern vary according to its nature (spatial or temporal). Highly representative and discriminative pattern features lead to a simplified classifier design. Irrelevant features must be discarded to enhance the system accuracy and performance. The classical pattern recognition systems often used a separate technique for feature extraction. The most popular feature extraction techniques are multidimensional scaling [49], linear discriminant analysis [50], principal component analysis [44], neural networks [44], fast transforms [51, 52], independent component analysis [53], moment invariants [54], genetic programming [55], evolutionary strategies [55], genetic algorithms [56–62], and fuzzified features [58–60].

For this research work, the features are cropped from entire facial images using Matlab software (Fig. 5). The features selected for this study are forehead, nose, right eye, left eye, mouth, and chin.

7. EMPIRICAL RESULTS AND DISCUSSION

The major problems which the designer of a pattern classification system faces using the most recent trend of combined classifiers are:

1. the selection of the base model,
2. the efficiency in time and space,
3. the scalability of the recognition task, and
4. the selection of the relavant features of the pattern to be tested.

Based upon these problems, a number of experiments have been performed to explore the possibility of the best parameters selection for invariant face recognition system.

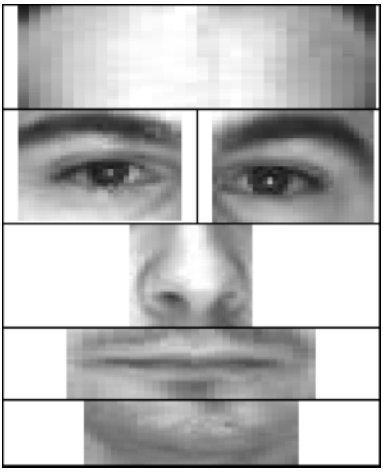

FIG. 5. Six extracted features of one person.

7.1. *Weak Classifiers*

The first set of experiments emphasizes the first aspect in the design of a pattern classification system, base model selection. This experiment uses a data set including 200 facial images of 20 persons. The training set includes $20 \times 5 = 100$ faces. The other 100 independent faces are kept away for test purpose. This experiment explores the effect of various training parameters on the performance of the LVQ base model such as:

1. the number of hidden layer neurons,
2. the initial weights for both the competitive and the linear layers,
3. the network building direction (growing up and pruning; elimination of redundant neurons), and
4. the learning rate.

Building a model could be achieved using either of the two approaches:

1. Increasing the hidden layer by successive incrementation of the number of hidden layer neurons according to the progress of the classifier performance using the percentage of correct classification of both training and test sets.

2. Starting with a large number of hidden layer neurons, evaluating the efficiency of each hidden layer neuron in representing a specific face class, using the standard deviation of the internal image, which the neuron has built up during the coarse of training process. A right internal representation of an input face results in an image with high contrast, resulting in high standard deviation. This internally built image represents the weight vector of the hidden layer neuron. When a specific hidden neuron fails to win the representation of an input face, the resulting internal image is blurred and its standard deviation is very low. This idea could then be used after each training epoch to eliminate *blind neurons* (those who fail in winning the representation of an input pattern). The killing process is performed by comparing the standard deviation for the

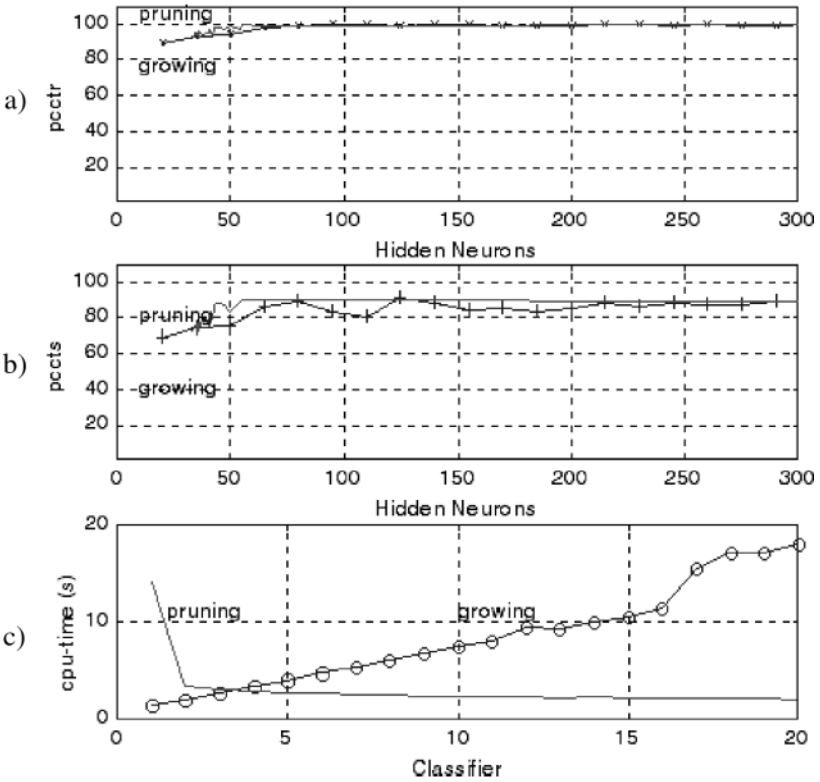

FIG. 6. Percentage of correct classification (pcc) and CPU time of growing up and pruning of hidden layer neurons networks for a set of 20 persons and five faces per person.

internal image with a specific threshold level. The threshold level is conducted experimentally for the underlying database.

In this experiment, the two previously described approaches are applied on a face database [22]. Figures 6a and 6b show the percentage of correct classificarion (pcc) rates (for cases of both training and test sets) for the two cases of hidden layer growing and hidden layer neuron pruning. For the case of hidden layer gowing, the initial number of hidden neurons is selected equal to 20 (minimum possible = number of target classes) and is increased with an incrementation step of 15. This process resulted in a total of 20 classifiers. For the case of the hidden layer neuron pruning scheme, the initial number of hidden layer neurons is 300 (20 persons $\times$ 3 representative neurons $\times$ 5 training faces per person). A number for eliminating neurons is fixed to 20 cycles to end up with 20 classifiers in the first case. Each classifier is trained in both cases for 5000 training epoches. From Figs. 6a and 6b we notice that the performance of the gowing classifiers is similar or inferior to that of the classifiers which rely on neuron elimination strategy. A second major advantage of neuron pruning is the faster training time. It could be noticed in Fig. 6c that the training time for pruning-based classifiers decays very fast compared with the growing classifiers which result in a linear increase in training time, while resulting in bad performance. Starting with 300 hidden neurons with a PCCTS of about 89% (PCCTR = 100%), the next classifier yields a PCCTS of 90% (PCCTR = 99%) with only 63 neurons. The variation in the learning rate also has the effect on the performance of the networks using pruning. It has been observed that the learning rate (α) of 0.7 yields the optimal performance in terms of the percentage of correct classification.

7.2. *Scalability Problem*

The procedure followed in the first experiment is repeated on larger databases including 500 faces (100 persons $\times$ 5 face images). The purpose of this experiment is to show that the developed algorithm is more efficient for larger databases than for smaller ones (Fig. 7). This advantage represents a major progress toward solving the scalability problem in pattern recognition tasks.

7.3. *Threshold Level*

This experiment is tried with various threshold levels for eliminating hidden layer redundant neurons (Table 2). It is inferred that the threshold level of 36 (standard deviation) generates the classifiers with better pcc performance for entire face images. This means that the threshold value varies for different objects depending upon the nature and size of the image. For the case of the

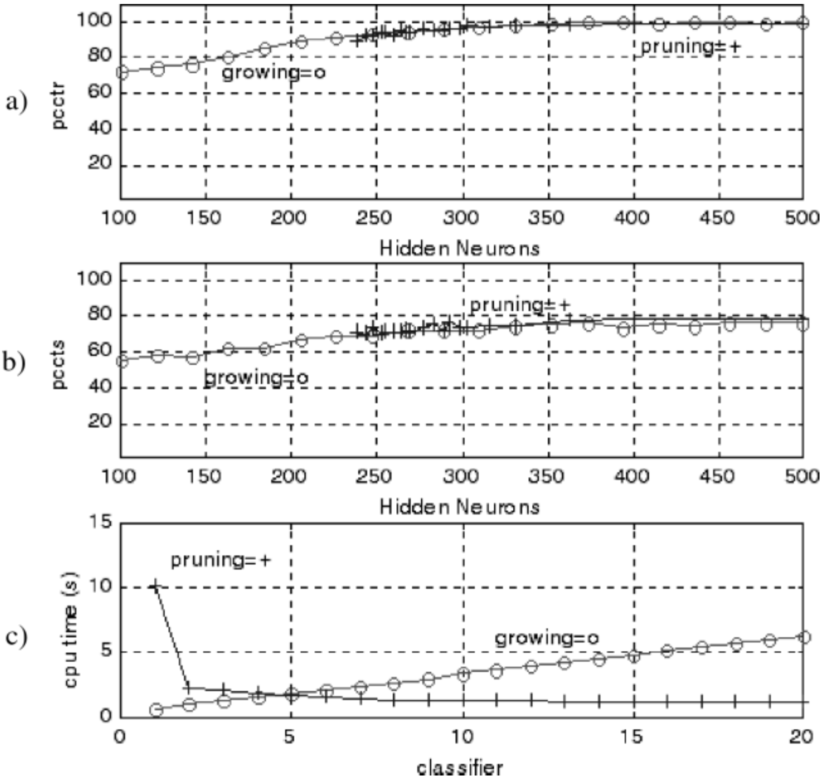

FIG. 7. Percentage of correct classification (pcc) and CPU time of growing up and pruning of hidden layer neurons networks for a set of 100 persons and five faces per person.

TABLE 2

Percentage of Correct Classification of the Test Sets (PCCTS) for Various Elimination Threshold Levels of Pruning and Growing Networks

CL	35	36	37	38	39	40	41	42	43	44	45	GN
1	87	85	88	88	85	87	88	87	85	88	89	87
2	90	82	87	87	82	87	86	91	82	87	90	88
3	89	93	86	85	87	86	87	86	87	85	90	88
4	86	90	85	85	85	81	86	81	85	83	89	87
5	84	86	82	82	80	80	88	80	80	86	83	81
6	88	87	79	79	81	77	87	80	75	81	87	85
7	87	85	84	84	82	80	77	76	78	83	88	86
8	85	87	82	82	77	76	79	81	76	82	87	85
9	92	90	80	81	82	76	81	78	81	80	79	77
10	88	88	77	77	77	74	83	79	77	79	71	69
11	86	90	74	74	71	77	85	73	71	73	75	73
12	89	89	83	83	78	77	80	80	78	75	79	77
13	85	89	72	72	70	80	75	82	70	72	71	69
14	86	91	81	81	74	71	72	82	74	73	73	71
15	88	91	76	76	77	76	76	79	77	76	75	73
16	83	86	81	81	71	81	78	75	71	74	73	71
17	88	84	76	75	73	79	76	81	73	71	74	72
18	84	89	76	76	72	78	79	72	72	72	71	69
19	88	86	83	83	76	73	72	80	76	76	76	74
20	85	91	82	82	77	79	76	75	77	75	72	70
AV	86.9	87.9	80.7	80.6	77.5	78.7	80.5	79.9	77.3	78.5	79.6	77.6

Note. CL, classifier; GN, growing network; AV, average.

nose, the experimental results show that the value of 25 (sdv = 25) yields the better pcc network performance. Also, it could be used as a variable that results in more efficient input classifiers to be combined. Figure 8 shows that the pool of models generated adopting the pruning network give a better percentage of correct classification for the test set (pccts) as compared to the growing network especially in the region of a low number of hidden layer neurons, i.e., classifiers 10 to 20 (Fig. 9).

7.4. *Image Size Independency*

The experiments are performed using various image sizes, 32×32, 48×48, and 64×64, for a training set of 20 persons and five faces per person. The network is tested on another set of 100 facial images. Figure 10 shows the relative pcc performace of both the pruning and the growing networks.

7.5. *Features Selection*

Features selection is an important problem when designing a pattern recognition system that is concerned with which attributes are most relevant for decision making. Feature selection plays a vital role in specifying the performance of the pattern classifier due to the following reasons:

1. redundant features can degrade the system performance,
2. improve the reliability of the estimate of performance,
3. more features mean higher feature extraction cost,

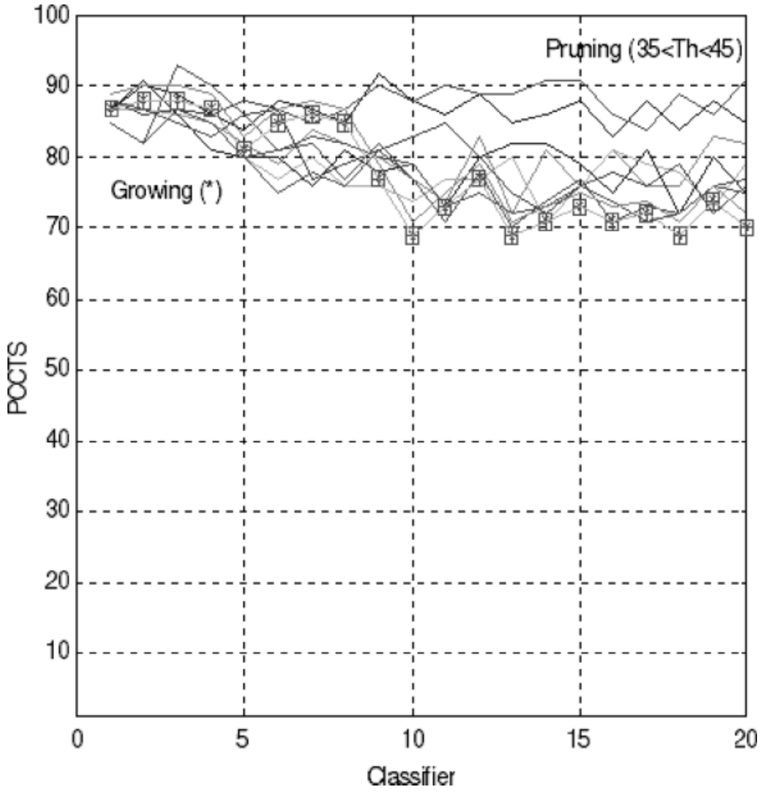

FIG. 8. Percentage of correct classification of test sets (pccts) vs classifiers for both pruning and growing networks.

4. reduce the training time in neural classifiers, and
5. avoid the curse of dimensionality.

As previously mentioned, the features selected for the study of this research work are forehead, nose, right eye, left eye, mouth, and chin (see Fig. 5). The output–decision vectors are then combined using majority voting. Also, the ranking of the class labels is used where the majority of the feature classifiers do not agree upon the same output class. The experiment is performed using a training set of 50 persons $\times$ 5 facial images $=$ 250 images for each of the features selected, i.e., a total of 1500 feature images for the training set and another set of the same number of images is used for testing out the network performance. It has been found experimentally that the recognition performance of the network for the forehead is the worst while the nose yields the best performance among the selected features. The ranking of the features in accordance with the pcc performance is nose, right eye, mouth, left eye, chin, and forehead. It may be noted that the recognition performance between left and right eyes is different. This is due to the database selected for the study. The chosen first five images of each individual are (1) neutral expression, (2) smile, (3) anger, (4) scream, and (5) left light on. Hence, the left eye is exposed to the light in the fifth pose–image of each person. This means that the selection of features for the face recognition system to be designed highly depends on the nature of data to be tested on and

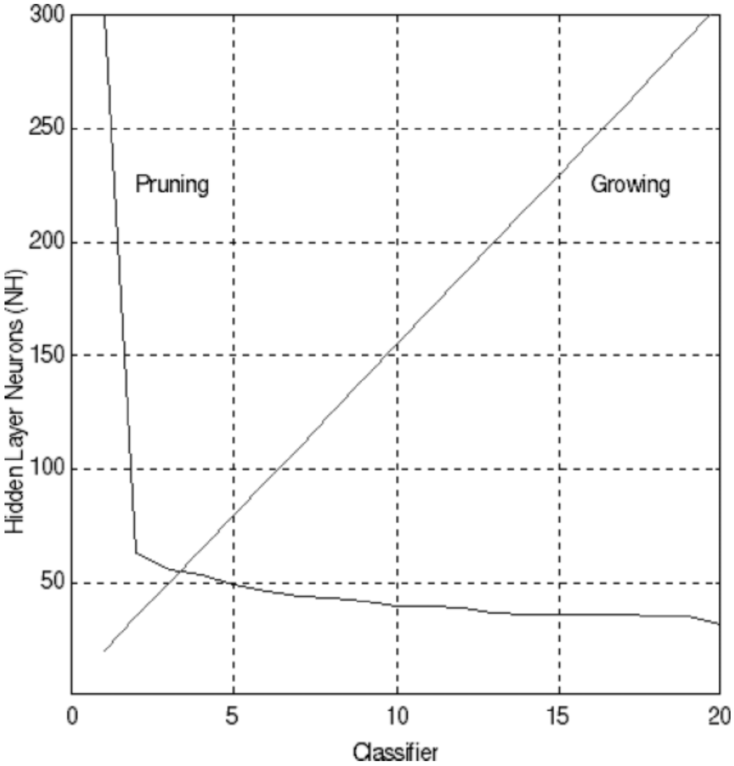

FIG. 9. Number of hidden layer neurons (NH) vs classifiers for both pruning and growing networks.

the feature region itself. Commonly, it is considered that the mouth and eyes are dynamic features as compared to the chin or forehead. However, experimental results show that the percentage of the correct classification rate for eyes is better than the chin or forehead, which are static features. This also means that when one region of the face is affected by the variation of the pose or expression, the other face regions are still unaffected. Thus, the recognition performance is high for the systems based on feature combination.

The simulation results show that the system is capable of identifying the face with 96.11% recognition rate trained with six facial features for 10,000 iterations. The recognition rate improves if the system is fed with the entire face in addition to the features and the decision is combined. The recognition rate of the system combined with the face and six features is 98.03%. For the case of 20 persons and 5 faces per person, the recognition rate of the system trained for 6000 iterations with or without combining the entire face is 98.67%. These results are the average of five runs. Also, the recognition rate could be improved using one of the well-known feature extraction techniques.

8. CONCLUSIONS

Invariant face recognition is a challenging task in computer vision. In classification tasks it may be wise to combine observations from different

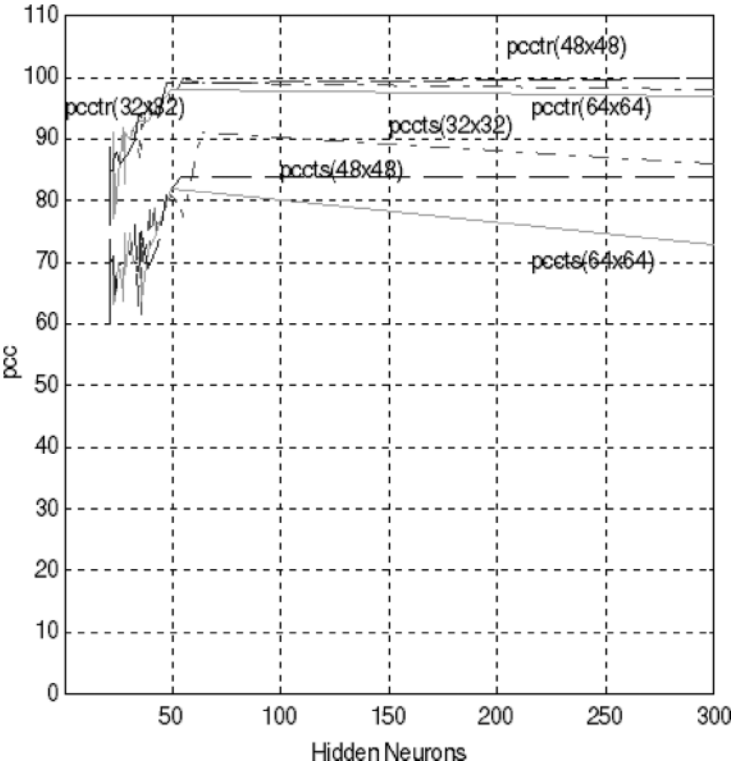

FIG. 10. Percentage of correct classification (pcc) stability @ 5000 iterations for various input image sizes.

sources. Not only does it decrease the training time but it can also increase the robustness and the performance of the classification. Combinations of weak classifiers, which use a combination scheme and a guided search (not random) algorithm, have shown the potential to achieve time efficiency as well as a good generalization performance. The combined classifiers show a good scaling property, which indicates efficiency in space complexity. The experiments are performed with various elimination threshold levels for the hidden layer neurons. It is shown that the threshold level could be used as a variable, which results in more efficient classifiers to be combined. Also, the network performance of the classifiers is shown to be independent of the size of images.

In modern pattern recognition systems all the stages of pattern recognition could be performed by a single scheme such as neural networks and genetic algorithms which has the inherent capabilities of noise filtering, data reduction, feature extraction, and classification. The advantage of using neural networks and genetic algorithms is that they can extract the most discriminative and representative set of features. In this case the input pattern to the neural network is the whole pattern (holistic approach). To avoid the long training time, data reduction through a feature extraction technique could be used.

To obtain classification systems with both good generalization performance and efficiency in space and time, a learning method based on combinations of weak classifiers is proposed. The weak classifiers are generated using automatic elimination of redundant hidden layer neuron networks on both the entire face images and the extracted features: forehead, right eye, left eye, nose, mouth, and chin. The classifiers are then combined through majority voting and ranking level on the decisions available from input classifiers. It is demonstrated that the proposed system is able to obtain better classification results with both good generalization performance and a fast training time on a variety of test problems.

Feature selection plays a vital role in specifying the performance of the pattern classifier. Highly representative and discriminative pattern features lead to a simplified classifier design. Irrelevant features must be discarded to enhance the system accuracy and performance. However, the selection of stable and representative sets of features that efficiently discriminate between faces in a huge database is the major problem. It can be concluded that the variability of facial pose, expressions, and lighting conditions renders it very difficult to rely on one set of features for all pattern recognition systems. The selection of features for the face recognition system to be designed highly depends on the nature of data to be tested on and the feature region itself.

Areas for future research include selection of a suitable set of features for generalized classification, investigation of better methods for classifiers combination, and expansion of the database.

ACKNOWLEDGMENTS

This work was supported by Kuwait University. The author is thankful to A. M. Martinez and R. Benavente for compiling and maintaining the AR Face Database.

REFERENCES

1. Baum, E. and Haussler, D., What size net gives valid generalization? *Neural Comput.* **1** (1989), 151–160.
2. Ji, C. and Psaltis, D., The VC-dimension versus the statistical capacity for two layer network with binary weights. In *Proc. 4th Workshop Comput. Learning Theory,* 1991.
3. Nowlan, S. J. and Hinton, G. E., Simplifying neural networks by soft weight sharing. *Neural Comput.* **4** (1992), 473–493.
4. Tishby, N., Levin, E., and Solla, S., Consistent inference of probability in layered networks: Prediction and generalization. In *Proc. Int. Joint Conf. Neural Networks,* 1989.
5. Valient, L. G., A theory of learnable. *Commun. Assoc. Comput. Mach.* **27** (1984), 1134–1142.
6. Vapnik, V., *Estimation of Dependences Based on Empirical Data.* Springer-Verlag, New York, 1982.
7. Barron, A., Approximation and estimation bounds for artificial neural networks. *Mach. Learning* **14** (1994), 113–143.
8. Barron, A., Predicted squared error: A criterion for automatic model selection. In *Self-Organizing Methods in Modeling* (Farlow, S. J., Ed.). Marcel Dekker, New York, 1984.
9. Devroye, L., Automatic pattern recognition—A study of the probability of error. *IEEE Trans. Pattern Anal. Mach. Intell.* **10** (1988), 530–543.
10. Geman, S., Bienenstock, E., and Doursat, R., Neural networks and the bias variance dilemma. *Neural Comput.* **4** (1992), 1–58.
11. Ho, T. K., Hull, J. J., and Srihari, S. N., Decision combination in multiple classifier systems. *IEEE Trans. Pattern Anal. Mach. Intell.* **10** (1994), 66–75.
12. Kleinberg, E. M., Stochastic discrimination. *Ann. Math. Artificial Intell.* **1** (1990), 207–239.

13. Xu, L., Krzyzak, A., and Suen, C. Y., Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Trans. Systems Man, Cybernet.* **22** (1992), 418–435.
14. Hashem, S., Optimal linear combinations of neural networks. *Neural Networks* (1994).
15. Tanigushi, M. and Tresp, V., Averaging regularized estimators. *Neural Comput.* **9** (1997), 1163–1178.
16. Tax, D. M. J., van Breukelen, M., Duin, R. P. W., and Kittler, J., Combining multiple classifiers by averaging or by multiplying? *Pattern Recog.* **33** (2000), 1475–1485.
17. Xu, L., Kryzak, A., and Suen, C. V., Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Trans. Systems, Man Cybernet.* **22** (1992), 418–435.
18. Battiti, R. and Colla, A. M., Democracy in neural nets: voting schemes for classification. *Neural Networks* **7** (1994), 691–707.
19. Tumer, K. and Ghosh, J., Order statistics combiners for neural classifiers. In *Proceedings of the World Congress on Neural Networks.* INNS Press, Washington DC, 1995, I: 31–34.
20. Jacobs, R., Method for combining experts' probability assessments. *Neural Comput.* **7** (1995), 867–888.
21. Rogova, G., Combining the results of several neural network classifiers. *Neural Networks* **7** (1994), 777–781.
22. Martinez, A. M. and Benavente, R., The AR face database. CVC Technical Report 24, June 1998.
23. Chellappa, R., Wilson, C. L., and Sirohey, S., Human and machine recognition of faces: A survey. *Proc. IEEE* **83** (1995), 705–740.
24. Valentin, D., Abdi, H., Toole, A. J. O., and Cottrell, W., Connectionist models of face processing: A survey. *Pattern Recog.* **27** (1994), 1209–1230.
25. Samal, A. and Iyangar, P., Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern Recog.* **25** (1992), 65–77.
26. Turk, M. and Pentland, A., Eigenfaces for recognition. *J. Cog. Neurosc.* **3** (1991), 71–86.
27. Belhumeur, P. N., Hespanha, J. P., and Kriegman, D. J., Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **19** (1997), 711–720.
28. Samaria, A. S., *Face Recognition Using Hidden Markov Models.* Ph.D. Dissertation, Trinity College, Cambridge University, Cambridge, UK, 1994.
29. Lawrence, S., Lee, G. C., Tsoi, A. C., and Back, A. A., Face recognition: A convolutional neural network approach. *IEEE Trans. Neural Networks* **8** (1997), 98–113.
30. Lin, S. H., Kung, S. Y., and Lin, L. J., Face recognition/detection by probabilistic decision-based neural network. *IEEE Trans. Neural Networks* **8** (1997), 114–132.
31. Feitosa, R. Q., Thomaz, C. E., and Veigo, A., Comparing the performance of the discriminant analysis and RBF neural network for face recognition. In *Proceedings of the 5th International Conference on Information Systems Analysis and Synthesis,* 1999, Vol. 6, pp. 50–56.
32. Srinivas, G. and Wechsler, H., Face recognition using hybrid classifiers. *Pattern Recog.* **30** (1997), 539–553.
33. Zhang, M. and Flucher, J., Face recognition using artificial neural network group-based adaptive tolerance (GAT) trees. *IEEE Trans. Neural Networks* **7** (1996), 555–567.
34. Chen, L. H., Chen, J. R., Liang, D., Deng, S. H., and Liao, H. Y., A minimum classification error method for face recognition. *Image Processing and Its Applications. Seventh International Conference, Conf. Publ.* 465, **2** (1999), 630–633.
35. Hagen, S., *Face Recognition—A Novel Technique.* Master thesis, University of Heidelberg, Germany, 1995.
36. Ben-Arie, J. and Nandy, D. A., Volumetric/iconic frequency domain representation for objects with application for pose invariant face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **20** (1998), 449–457.
37. Kin-Man, L. and Hong, Y., An analytic-to-holistic approach for face recognition based on a single frontal view. *IEEE Trans. Pattern Anal. Mach. Intell.* **20** (1998), 673–686.
38. Li, S. Z. and Lu, J., Face recognition using the nearest feature line method. *IEEE Trans. Neural Networks* **10** (1999), 439–443.
39. Sutherland, K., Renshaw, D., and Denyer, B. P., A novel automatic face recognition algorithm employing vector quantization. In *Proc. IEE Colloquium on Machine Storage and Recognition of Faces,* 1992, pp. 41–44.

40. Wiskott, L., Fellous, J. M., Krueger, N., and von der Malsburg, C., Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **19** (1997), 775–779.
41. Lin, C.-H. and Wu, J.-L., Automatic facial feature extraction by genetic algorithms. *IEEE Trans. Image Process.* **8** (1999), 834–845.
42. Yoon, C., Park, J., and Park, M., Face recognition using wavelets and fuzzy C-means clustering. In *TENCON 99. Proceedings of the IEEE Region 10 Conference*.
43. Kohonen, T., *Self-Organizing Maps.* Springer-Verlag, Berlin, Germany, 1995.
44. Haykin, S., *Neural Networks: A Comprehensive Foundation,* 2nd ed., Prentice Hall, Englewood Cliffs, NJ, 1999.
45. Tumer, K. and Ghosh, J., Theoretical foundations of linear and order statistics combiners for neural pattern classifiers. Technical Report 95-02-98, Computer and Vision Research Center, University of Texas, Austin, 1995.
46. Ma, S. and Ji, C., Performance and efficiency: recent advances in supervised learning. *Proc. IEEE* **87** (1999).
47. Keller, P. E., Kangas, L. J., Liden, L. H., Hashem, S., and Kouzes, R. T., Electronic noses and their applications. In *IEEE Technical Applications Conference (TAC'95) at Northcon'95: Neural Network Application Studies Workshop,* 1995, pp. 116–119.
48. Hassoun, M. H., *Fundamentals of Artificial Neural Networks.* MIT Press, England, 1995.
49. Van Tasell, D. J., Speech waveform envelope cues for consonant recognition. *J. Acoust. Soc. Am.* **82** (1987), 1152–1161.
50. Feitosa, R. Q., Thomaz, C. E., and Veigo, A., Comparing the performance of the discriminant analysis and RBF neural network for face recognition. In *Proceedings of the 5th International Conference on Information Systems Analysis and Synthesis,* 1999, Vol. 6, pp. 50–56.
51. Gonzalez, R. C. and Woods, R. E. *Digital Image Processing.* Addison Wesley, Reading, MA, 1993.
52. Pittner, S. and Kamarthi, S. V., Feature extraction from wavelet coefficients for pattern recognition tasks. *IEEE Trans. Pattern Anal. Mach. Intell.* **21** (1999), 83–88.
53. Hyvaerinen, A., Survey on independent component analysis. *Neural Comput. Surv.* **2** (1999), 94–128. Available at http://www.icsi.berkley.edu/~jagota/NCS.
54. Pitas, I., *Digital Image Processing Algorithms,* Prentice Hall, New York, 1993.
55. Bäck, T., Hammel, U., and Schwefel, H.-P., Evolutionary computation: Comments on the history and current state. *IEEE Trans. Evol. Comput.* **1** (1997), 3–17.
56. Goldberg, D. E., *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley, Reading, MA, 1989.
57. Yow, K. C. and Cipolla, R., Enhancing human face detection using motion and active contours. In *Proceedings of the 3rd Asian Conference on Computer Vision (ACCV'98), Hong Kong, China,* 1998.
58. Lazzerini, B. and Marcelloni, F., A linguistic fuzzy recogniser of off-line handwritten characters. *Pattern Recog. Lett.* **21** (2000), 319–327.
59. Ramze, M., Goedhart, B., Leliveldt, B. P. F., and Reiber, J. H. C., Fuzzy feature selection. *Pattern Recog.* **32** (1999), 2011–2019.
60. Rajat, K. D. E., Pal, N. R., and Pal, S. K., Feature analysis: Neural network and fuzzy set theoretic approaches. *Pattern Recog.* **30** (1997), 1579–1590.
61. Deng, P. S., Liao, H.-Y. M., Ho, C. W., and Tyan, H.-R., Wavelet-based off-line handwritten signature verification. *Comput. Vision Image Understanding* **76** (1999), 173–190.
62. Konen, W., Maurer, T., and von der Malsburg, C., A fast dynamic link matching algorithm for invariant pattern recognition. *Neural Networks* **7** (1994), 1019–1030.

G. A. KHUWAJA received his Bachelor of Engineering in electronic engineering from the Mehran University of Engineering and Technology, Pakistan, in 1988. He received his M.Sc. and Ph.D. in electrical engineering from the University of Manchester Institute of Science & Technology (UMIST), England, in 1990 and 1992, respectively. From 1993 to 1996, he was an assistant professor of computer engineering at the British Universities Programme, Pakistan. Since 1997, he has been with the Department of Physics, Kuwait University as an assistant professor. His research interests include digital systems design, CAD, modern logic design, Reed–Muller algebra, switching theory, neural networks, digital signal processing, and multimedia data compression. He has published 15 papers in these areas. Dr. Khuwaja was the recipient of The Heaviside Premium award for the best paper of the year 1992/93 sponsored by the IEE Council, UK. He is an associative member of the IEE (UK) and a member of PEC (Pak).