Pattern Analysis & Applications (2000)3:289–302

Combined Classifiers for Invariant Face Recognition

A. S. Tolba and A. N. Abu-Rezq

Department of Physics, Kuwait University, Safat, Kuwait

Abstract: This paper presents a system for invariant face recognition. A combined classifier uses the generalisation capabilities of both Learning Vector Quantisation (LVQ) and Radial Basis Function (RBF) neural networks to build a representative model of a face from a variety of training patterns with different poses, details and facial expressions. The combined generalisation error of the classifier is found to be lower than that of each individual classifier. A new face synthesis method is implemented for reducing the false acceptance rate and enhancing the rejection capability of the classifier. The system is capable of recognising a face in less than one second. The well-known ORL database is used for testing the combined classifier. Comparisons with several other systems show that our system compares favourably with the state-of-the-art systems. In the case of the ORL database, a correct recognition rate of 99.5% at 0.5% rejection rate is achieved.

Keywords: Classification; Combined classifiers; Invariant recognition; Face recognition; Learning vector quantisation; Radial basis function network

1. INTRODUCTION

Invariant face recognition is a challenging task, especially in the absence of highly controlled environments and recognition constraints. Recent progress of computer technology has made us expect that the face will play a key rule in future human-machine interaction and advanced communications, such as multimedia and low-bandwidth video-telephony. This work is concerned with the task of face classification only. Face detection is beyond the scope of this study. In Hoogenboom [1], Daugman [2], Rowley et al [3] and Sung and Poggio [4], excellent face detection results were reported. A detailed overview of face recognition approaches can be found in the extensive surveys by Chellapa et al [5], Valentin et al [6] and Samal and Iyanger [7]. Recent research focuses on pose invariant face recognition. In this paper, we present a neural network-based combined classifier for invariant recognition. Combining component classifiers into a composite classifier have been successfully used in various application fields such as handwritten character recognition [8], and face recognition [9,10]. Combined classifiers could be broadly categorised as:

Received: 26 May 1999
Received in revised form: 16 March 2000
Accepted: 16 March 2000

1. category-based classifiers (heterogeneous classifiers that come from different model classes), and
2. parameter-based classifiers (homogeneous classifiers that come from the same model classes).

The homogeneous classifiers could be trained on different feature vectors. Heterogeneous classifiers could be trained either on the same feature vectors or on different feature vectors.

The combined classifier results in better generalisation and higher accuracy than the most accurate component classifiers under the condition that the component classifiers are independent and unbiased. Selection of the component classifiers is not yet well known [11]. Most of the recent research work in this field tries to combine the most powerful components. An exception is the approach described in Chuanyi and Sheng [12], where a learning method based on combination of weak classifiers (linear perceptrons) is reported. The goal is to obtain classification systems with both good generalisation performance and efficiency in space and time. To select the component classifiers, a randomised algorithm is proposed. A majority vote is then used to combine the selected weak classifiers. Experiments demonstrated a combination of weak classifiers with good generalisation performance and a fast training time on a variety of test problems and real applications. The price paid to achieve this is a larger space-complexity compared to that of a well-trained classifier.

The classifier that is implemented in this paper is a category-based classifier that combines the generalisation characteristics of both the LVQ and RBF classifier networks. The problem which must be solved is that the two component classifiers always agree on the final decision during the test phase. This leads to a useless combination. To overcome this problem, the problem of separating familiar from non-familiar faces, will be addressed. Separating the training set into two sets of familiar and non-familiar faces, and training one of the classifiers on the first set and the second classifier on the second set, helped in improving the generalisation capability of the combined classifier. Generating new faces from combinations of the faces, which confuse the classifiers using a LVQ classifier, enhances the rejection performance of the individual classifiers.

In the following sections, we describe each component of the proposed system. Section 2 gives a short overview of the previous face recognition techniques, specifically those applied to the ORL database [13]. Section 3 described the ORL database used in this research, and the preprocessing of face images. Section 4 presents the individual classifiers (Learning Vector Quantisation classifier and Radial Basis Function classifier) and the combined classifier. The experimental results of both the individual and combined classifiers are presented and discussed in Section 5. In Section 6, a comparison is performed between the proposed system and the previously reported systems in the literature. Finally, the summary and conclusions are given in Section 7.

2. OVERVIEW

Face recognition approaches could be categorised into two major categories: feature-based approaches and holistic approaches. Figure 1 shows the different sub-categories of each approach. A detailed overview of face recognition approaches can be found in several extensive surveys [5–7]. In this section, the review is restricted to the discussion of those approaches that were applied to the ORL image database, which was developed at the Olivetti Research Laboratory in Cambridge [13].

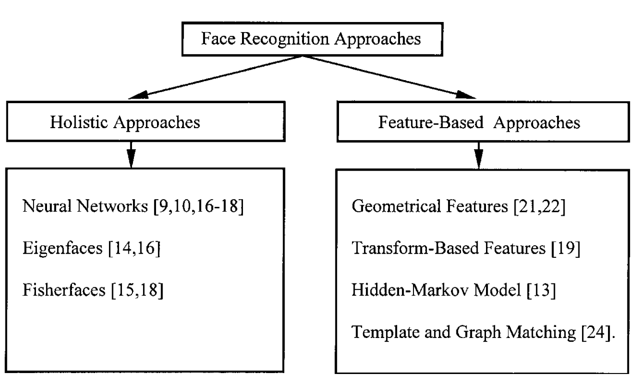

Fig. 1. Face recognition techniques.

2.1. Holistic Approaches for Face Recognition

The eigenface approach described by Turk and Pentland [14] 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 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 [14] reported a correct recognition rate of 95% in the case of the FERRET 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 is that the scatter being maximised 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 [15].

Many other researchers have implemented the eigenface approach for comparison purposes. Belhumeur et al [16] 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 maximise the 'between-class scatter' while minimising the 'within-class scatter'. This approach is proved to be more efficient than the eigenface approach, especially in the case of variable illumination. The experiments were performed on only 150 faces from 15 subjects selected from the ORL database. 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 [15], testing the eigenface method on the ORL database resulted in 89.5% correct recognition rate. Both a convolutional neural network and a self-organising 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 [17], a Probabilistic Decision-Based Neural Network (PDBNN) is described for face recognition. While the system performance in the case of the FERRET database is 99%, its performance for the case of the ORL database is 94%. The face recognition time is less than 0.1 seconds on an SGI Indy machine.

In Feitosa et al [18], the performance of both the Linear Discriminant Analysis (LDS) 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 (five images × 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. The Principal Component Analysis (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 generalisation 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 [9], 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 [10], a Group-based Adaptive Tolerance (GAT) 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 large number of classes and noisy inputs, and is capable of being upgraded to recognise new tasks without the need for retraining. The face recognition system locates the captured faces using MLP neural networks. A GAT tree is used in the middle level for face recognition, using normalised facial images. The face is classified either 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 grey 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 recognise titled and rotated faces with similar confidence levels. Training is performed on 136 faces, and testing with 653 faces. The observed errors for tilted and rotated face recognition were 0.16% and 0.31%, respectively. To recognise front beard faces, six faces were chosen to train each GAT node (three front beard faces, three non-beard 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 tree to noise was tested, and is found to be more than that of the general trees.

2.2. Feature-Based Approaches for Face Recognition

In Samaria [13], 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 characterised 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 [19], 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 together with the Euclidean classifier with both the Back-Propagation (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 [20], 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 recognised in 320 seconds.

In Kin-Man and Hong [21], 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 pre-stored features. In addition to that, 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 [22], 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 generalise 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%, and the authors claim that it is the lowest reported rate for the ORL database. 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 [23], 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 Quantisation (VQ) of the facial features is performed after these features have been extracted from the entire image. One vector quantiser is dedicated to each of the eight features used. The vector quantisers 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 sub-sampled due to their relative unimportance in frontal face recognition. An additional image of each person was manually segmented and used to form the vector quantiser 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 inter-relationships has not been integrated in the VQ coefficient analysis. The test results for 300 test images were 89.19%. The VQ technique described above suffers from the following deficiencies:

1. Manual intervention for preparation of eight training facial features.
2. Elimination of the inter-relations between the eight facial features.
3. Only frontal faces were considered in the database.

In Wiskott et al [24], 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 (labelled 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 grey level features. For image matching against a stored graph, the graph location in the image is optimised. It has been shown that Elastic Graph Matching can successfully recognise faces from facial line drawings. The efficiency of the Gabor-wavelets in recognising 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 classifications experiments performed on line drawings resulted in a correct-decision rate of better than 90%.

2.3. Summary and Discussion of Previous Work

Table 1 summarises the results of the previously described approaches. The previous overview leads to the following conclusions:

1. Early research in the face recognition field concentrated on the extraction of geometrical features for a description of the shape of facial components like the mouth, eyes and muscle motion [5]. Since faces are dynamic objects that undergo a vast number of non-rigid deformations, which may vary from one person to the other, geometrical features are not a suitable measure for a well-defined description of such deformations.
2. Although eigenfaces provide a compact representation of whole faces, they do not provide invariance over changes in scale, head orientation and lighting conditions [15,25]. The internal representations in the auto-associative Multi-Layer Perceptron (MLP) are essentially equivalent to the principal components [26, p. 316]. However, in the case of feed-forward neural networks, the face subspace is identified according to the training stage and is appli-

Table 1. Summary of face recognition approaches when applied to the ORL database

Approach	Reference and notes	PCR (%)
Eigenfaces	Lawrence et al [15]	89.5
	Hagen [19]	94.0
	Wei et al [25]	95.0
Fisherfaces	Feitosa et al [18]	95.7
Feature-Based	Ben-Arie et al [20] (Geometrical features)	92.5
	Kim-Man et al [21] (Geometrical features)	84.0
	Hagen [19] (2D-FTT + Nearest Neighbour Classifier)	98.0
Graph Matching	Li et al [22] (Nearest Feature Line)	96.8
Hidden Markov Model	Samaria [13] (Hidden Markov Model)	95.0
Neural Networks	Hagen [19] (MLP)	96.5
	Lin et al [17] (PDBNN)	94.0
	Lawrence et al [15] (SOM + CNN)	96.2
	Tolba and Abu-Rezq (this paper) (LVQ)	99.0
	Tolba and Abu-Rezq (this paper) (RBF)	98.0
	Tolba and Abu-Rezq (this paper) (Combined LVQ, RBF)	99.5

cation specific. The relationship between the internal representation of knowledge in neural networks and the eigenvector decomposition was studied by several authors [27,28], who have proved that the neural network learning can approximate an eigenvector description of the data presented. In Valentin and Abdi [27], it has been shown that using the auto-associative memory model to store and retrieve faces is equivalent to computing the eigen-decomposition of the set of the faces as a weighted sum of eigenvectors.

3. Radial Basis Function Network [18], Convolutional Networks and Self-Organising Maps [15] were not extensively studied to explore their actual generalisation capability in the context of invariant face recognition. Although the RBF networks are computationally favourable compared to the other neural networks, the results reported to-date are not satisfactory [18].
4. The face recognition approaches such as eigenfaces [14] and PDBNN [17], which resulted in high recognition rates (95% and 99%, respectively) when applied to the FERRET database (3000 faces), result in inferior performance (89.5–95.5% and 94%, respectively) when applied to the ORL database.
5. In almost all cases, carefully designed neural network classifiers [15,17] which are trained on raw image pixels result in superior performance when compared to geometrical feature based methods [20,21].

3. DATABASE AND PREPROCESSING

The investigations described in this paper were performed using facial images of the ORL database. The ORL image database [14] was developed at the Olivetti Research Laboratory in Cambridge. The data consists of 400 images acquired from 40 people, some of which were taken at different times for some of the people. There are variations in facial expression (open/closed eyes, smiling/non-smiling), and facial details (glasses/no glasses). All images were taken against a dark homogenous background with the subjects in an upright frontal position, with a tolerance for some tilting and rotation of up to 20 degrees. There is some variation of the scale of up to about 10%. The images are grey scale with a resolution of 92×112 pixels. The images are size normalised. Figure 2 shows the faces of 40 individuals included in the ORL database. The whole set of images is re-sampled to three different sizes: 24×24, 32×32 and 64×64. To reduce the image size, a low pass filter is applied to the image before interpolation using the nearest neighbour interpolation method. This reduces the effect of Moiré patterns and ripple patterns that result from aliasing during re-sampling. After re-sampling, all images will have the same size.

4. NEURAL CLASSIFIERS

Although better than the feature-based approaches, almost all of the previous applications of neural networks to face recognition resulted in high error rates. Both the convolutional neural network combined with the self-organising map [15] and the probabilistic Decision-based Neural Network (PDBN) [17] resulted in high error rates (3.8% for the former and 6% for the latter) when applied to the ORL database. In Hagen [19], the MLP neural network resulted in an error rate of 4%. In this section, we present the application of two well-known neural networks (LVQ and RBF) to the problem of invariant face recognition. The LVQ aims at defining decision surfaces between the competing classes. The decision surfaces obtained by a supervised stochastic learning process of the training data are piecewise-linear hyper-planes that approximate the Bayesian minimum classification error probability.

It is well-known that the Radial Basis Function Networks (RBFN) overcome some of the MLP problems by relying on a rapid training phase, and presenting systematic low responses to input patterns that have fallen into regions of the input space where there are no training samples. Such characteristics and the intrinsic simplicity of these networks render RBFN classifiers an interesting alternative to classifiers based on other neural models [27]. However, the classification error made by the RBFN classifiers strongly depends

Fig. 2. Samples of faces of the 40 people included in ORL database [13].

upon the selection of the centres and widths of the kernel functions associated with the hidden neurons if the network. Selection of the suitable design-parameters of both classifiers is conducted experimentally after performing an extensive set of experiments. An iterative technique is used for the effective selection of the best spread constant of the kernel function on the basis of the best correct recognition rate.

4.1. Learning Vector Quantisation (LVQ) Classifier

Learning vector quantisation is a supervised classifier that was first studied by Teuvo Kohonen [29]. Several variations on the basic LVQ algorithm have been proposed by Kohonen. 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 has been demonstrated to coincide closely with those of a Bayes classifier. Each region is represented by a prototype vector. 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. 3) consists of an input layer, a hidden competitive layer, which learns to classify input vectors into subclasses, and an output linear layer, which transforms the competitive layer's classes into target classifications defined by the user. Only the winning neuron of the linear 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

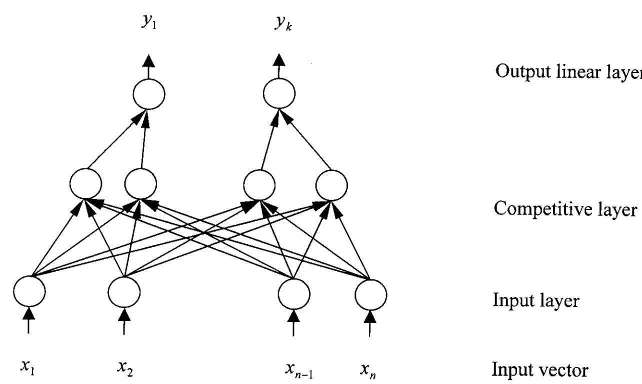

Fig. 3. Architecture of the LVQ classifier.

relationship, and significantly affects the results of classifier testing. Selection of the number of hidden neurons must be carefully made, as it highly depends upon the encompassed variability in the input patterns. Extensive experiments are performed to conduct the suitable number.

In the case of a training set consisting of n input faces, each of these faces is labelled as being one of k classes. The learning phase starts by initialing the weight vectors of the neurons in the hidden layer. Then, the input vectors are presented randomly to the network. For each input vector x_i, a winner neuron w_i is chosen to adjust its weight vector:

$$\|x_i - w_i\| \leq \|x_i - x_k\| \quad \text{for all } k \tag{1}$$

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

$$w_i(t+1) = w_i(t) + \alpha(x_i - w_i(t)) \tag{2}$$

if x_i and $w_i(t)$ belong to the same class

$$w_i(t+1) = w_i(t) - \alpha(x_i\, w_i(t)) \tag{3}$$

if x_i and w_i belong to different classes

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 [30]. Otherwise, do not change the weights. The training algorithm is stopped after reaching a pre-specified mean-squared error limit. During the testing phase, the distance of an input vector to each processing element of the hidden layer is computed, and again the nearest element is declared as the winner. This, in turn, fires one output neuron, signifying a particular class.

4.2. Radial Basis Function (RBF) Classifier

In theory, the RBFN, like the Multi Layer Perceptron (MLP) is capable of forming an arbitrary close approximation to any continuous nonlinear mapping. The RBF network has been identified as valuable model by a wide range of researchers [26,31,34]. The main advantages of a RBF network are the computational simplicity and its description by a well-developed mathematical theory. Radial Basis Functions can be designed in a fraction of the time it takes to train a standard feed-forward network. They work best when many training vectors are available.

An RBF network (Fig. 4) consists of three layers: the input layer, which has a number of units equal to the dimension n of the input vector; the hidden layer, which contains a number of RBF neurons that is equal to the number of training patterns included in the database; and the output linear layer, which has a number of units m that depends upon the number of classes of interest. Initially, the weight vector between an input unit i and the jth RBF neuron is simply equal to the input vector of the jth sample of the training set: $w_j = x^j$. The output of the ith neuron of the output layer is then

$$y_i(x) = \sum_{j=1}^{N} w_{ij}\phi\, (\|x - x^j\|) \tag{4}$$

where $\phi(\cdot)$ is a decreasing function, x is the input vector,

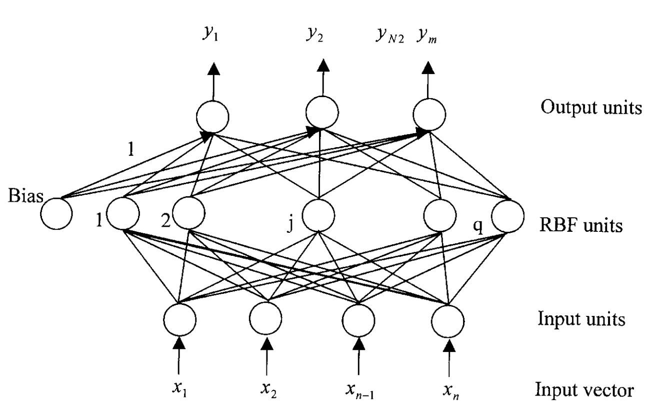

Fig. 4. Architecture of the RBFN classifier.

and x^j are input examples of the training set. The rule of the training phase is to tune the weights w_{ij} between the RBF units and the input units. The transfer function of the radial basis function neuron is the Gaussian function

$$G(x,x^j) = \exp\left(-\frac{1}{2\sigma_j^2}\|x - x^j\|^2\right) \tag{5}$$

The RBF network approximation becomes

$$f(x) = \sum_{j=1}^{N} w_j \exp\left(-\frac{1}{2\sigma_j^2}\|x - x^j\|^2\right) \tag{6}$$

Chen et al [32] proposed an Orthogonal Least-Squares (OLS) supervised-algorithm to select, one by one, the best centres c_j within training samples x^i, $1 \leq i \leq q$, as centres:

$$y(x) = \sum_{j=1}^{q} w_j\phi(\|x - x^j\|) \tag{7}$$

A RBFN classifier is trained using a set of input vectors that represent the lexicographically ordered rows of the facial images in the training set, the vector of target classes, and a spread constant for the radial basis layers. The spread constant is the only parameter that has to be adjusted to ensure that the active regions of the radial basis neurons overlap enough so that several neurons have fairly large outputs at any given moment. This makes the network function smoother, and results in better generalisation for new input stimuli occurring between input stimuli used in the training phase.

Training starts by creating one radial basis neuron at a time. The addition of neurons continues until the sum-squared error falls beneath an error goal or a maximum number of neurons has been reached. The input vectors, which will result in reducing the network error the most, are used to create a radial basis neuron. The error is checked after each-iteration, and if low enough, training is finished. The spread parameter determines the width of an area in the input space to which each neuron responds. Biases are set as a function of the spread, such that each neuron covers a specific area of the input space. The weights and biases of the linear layer are calculated such that the minimum sum-squared error goal is reached.

4.3. Combined Classifiers

In fact, a growing body of literature [11,12,33] has indicated that combining a set of classifiers is an effective way of improving expected generalisation performance. The best conditions for combining occur when the learned models are fairly accurate, but fairly independent in the errors they make. This can be achieved by using different feature vectors, different training sets, different training parameters or different classifier architectures.

The three criteria which must be fulfilled during the design of a combined classifier are accuracy of component classifiers, diversity of component classifiers, and efficiency of the entire composite classifier.

Many experiments were performed on the individual clas-

P100

P171

P200

Fig. 5. Familiar faces (misclassified by individual classifiers: LVQ and RBF).

sifiers to obtain the best performer (Section 5). Component accuracy could be sacrificed in favour of increasing accuracy [11]. As will be seen in the next section, although different, the two individual classifiers (LVQ, RBF networks) virtually agree upon the final decision of classifying the faces in the ORL database. This means that the diversity criterion is not satisfied because both classifiers misclassified the test patterns P100, P171 and P200 in the left-most column of Fig. 5. The faces to the right of the left-most column belong to the classes with which the misclassified patterns interfere. This situation means that these patterns have special attributes, which cause their interference with other classes. The classes of these patterns, together with the interfering classes, could be designated as Familiar Classes (FC). The other correctly recognised classes are designated as Distinctive Classes (DC). Considering the above diversity problem, designing a special type of classifier (Fig. 6) is necessary to resolve the confusion problems caused by the familiar faces. The classifier design steps are:

1. Train the best performing individual classifier (LVQ) on the whole training set (200 faces from 40 people).
2. Test the LVQ classifier of step 1 on the whole test set (other 200 faces from 40 persons).
3. Separate the training set into two groups: the correctly classified group (DC), and the misclassified group, together with the classes with which they interfere (FC).
4. Train a new LVQ classifier on the DC faces and a new RBF classifier on the familiar class of faces (FC).

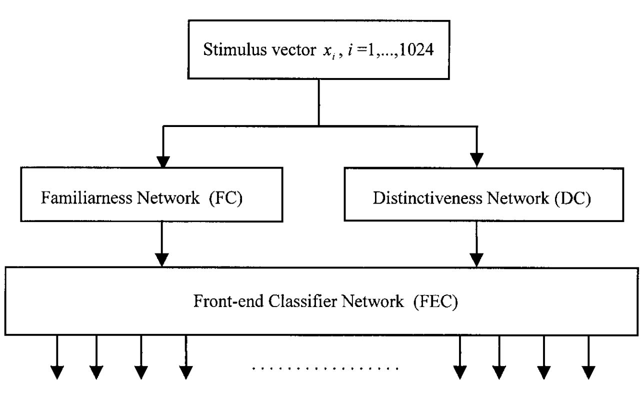

Fig. 6. Combined Face-Classifier (CFC).

5. Apply a Front-End Classifier (FEC) on the outputs of the DC and FC classifiers.

The efficiency criteria could be satisfied by avoiding prohibitively expensive classifiers such as the MLP. The RBF network and the LVQ classifiers are much faster than the MLP classifier.

A second problem which must be solved is the Error-Reject trade-off in each classifier. This problem is solved from the beginning for the FEC classifier if it is solved for the DC and FC classifiers. A reject class is considered during the training of the classifiers. Faces external to the database of interest, together with faces merged from faces included in the confused classes of the database, are included in the training set for the reject class. Figure 6 shows the architecture of the combined classifier.

5. EXPERIMENTS ON THE ORL DATABASE

In this section, three experiments were performed to evaluate the face recognition system. The first two preliminary experiments are performed on the two individual classifiers (LVQ and RBF) in order to select the best candidates for constructing the combined classifier. The third set is made to test the combined classifier. The investigated aspects of the experiments are:

1. Selection of the suitable neural network based classifier.
2. Selection of the number of training faces which could encompass the possible sources of variability such as different- poses, sizes, detail and facial expressions.
3. Selection of the optimal face image size for compromise between classifier correct rate and classification time.
4. Selection of the architectural parameters of the individual neural classifiers.
5. Selection of the optimal training parameters like the learning rate and number of training epochs.
6. Enhancing the rejection performance of the classifiers using synthetic faces generated by the LVQ classifier.
7. Testing of the individual and combined classifiers on an unseen set of 200 test faces.

5.1. Experiments on the LVQ Classifier

5.1.1. Selection of the LVQ Classifier Parameters. Images of the ORL database are categorised into two sets: a training set of 200 faces; and a test set of other unseen 200 faces (five faces × 40 people). The weighting matrices of both layers of the LVQ neural network are first initialised. The learning rate is also initialised to a value of 0.5. The user specifies the target vector including the 40 target classes. Different experiments were performed to study the effect of different Training Epochs (TE), numbers of competitive layer neurons (NH), image sizes, Learning Rate (LR) and the number of training faces per individual. Table 2 shows the effect of different aspects of system design on the Percentage Correct Recognition (PCR) rate.

Table 2 indicates the relative effect of the different parameters on the system's performance. Investigation of the experimental results lead to the following conclusions:

1. The number of hidden layer neurons significantly affects the classifier performance (experiment 1 in Table 2). The number of neurons (1200) in the hidden layers gives a better chance to the classifier for capturing all the possible sources of image variability.
2. The learning rate greatly affects the system performance, and depends upon the selected number of training epochs, image size and number of hidden layer neurons. Therefore, there is no rule for selecting a suitable learning rate (experiment 2 in Table 2). Experiments should be performed to conduct the suitable rate in the context of the suitable network architecture.
3. The image size (24 × 24) resulted in inferior system performance compared to the other sizes (32 × 32 and 64 × 64) sizes for the same learning rate (0.5) and number of training epochs (8000). The image size (32 × 32 pixels) is a good compromise between computational requirements and recognition accuracy (experiment 3 in Table 2).
4. The network is over-trained for the case of 12,000 training epochs; this means that the accuracy starts to decrease slowly (experiment 4 in Table 2). The possible explanation is that the prototype vectors become very specifically tuned to the training set.
5. The ability of the algorithm to generalise for new data suffers. Therefore, experiments have to be conducted to find the suitable number of training epochs, depending on the training parameters and network architecture.
6. At least five training faces/individual must be used to achieve acceptable performance level (experiment 5 in Table 2). Selection of both the number of training faces and their poses is the most critical factor in solving the problem of invariant face recognition. Both the training and testing sets should include representatives for the frontal face, left turned face and right turned face.

5.1.2. Face Synthesis Using LVQ and Real Faces. A neural network-based method is used to create new face images from real ones in order to increase the diversity of the database that is used for training a classifier, which can then result in a significant improvement of the generalisation ability of the recognition system. This approach is particularly suitable, as the classifier is based on an artificial neural network, such as an LVQ classifier, which is able to assimilate the variations of the faces during its training phase. The face generation mechanism that is inherent in the structure of the designed LVQ network (Fig. 7(a): 32 × 32 input neurons, 1 hidden neuron and 1 output neuron) is based only on real facial images. Moreover, the generation of new images (Fig. 7(b)) from original faces only depends upon a small number of parameters. This offers the advantage that the parameters of the generating algorithm are very easy to set out, so as to produce images that are different enough from the original ones to bring new useful information, as well as to avoid creation of over-noisy images. The target for training the LVQ network on both faces is set to 1, which forces the network to combine the facial features of both images in the training set. The resulting faces of Fig. 7(b) indicate that when a neuron responds to two or more facial images, it generates a blend of these images, resulting in:

1. A blurred face when the input images nearly have the same view.
2. A face with multiple noses, eyes or mouths for input images with different poses.

5.1.3. Visualisation of the Weights of the Competitive Layer Neurons. The hidden layer neurons in a neural network, which is trained on a limited set of facial images, try to generalise to all the possible variations among the

Table 2. Performance of the LVQ classifier on a test set including 200 faces

Experiment 1		Experiment 2		Experiment 3		Experiment 4		Experiment 5	
NH	PCR	LR/TE/Size	PCR	Size	PCR	TE	PCR	Faces	PCR
200	96.0	0.05/8000/32 × 32	95.5	24 × 24	93.5	3000	92.5	4	95
400	97.5	0.50/8000/32 × 32	97.5	32 × 32	**97.5**	4000	95.5	5	99
800	98.0	0.05/8000/64 × 64	97.0	64 × 64	96.5	6000	96.0	6	**100**
1200	**99**	0.50/8000/64 × 64	96.5			8000	**97.5**		
						12000	94.5		

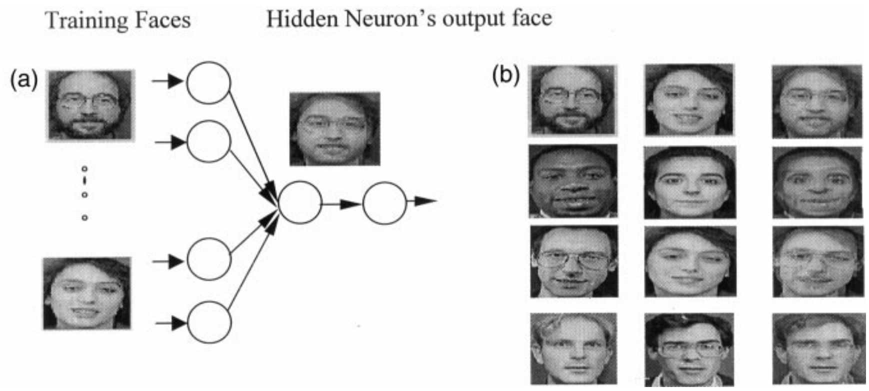

Fig. 7. (a) LVQ face synthesiser; (b) examples of face merging using LVQ.

whole set of faces so that each face has a representative for these variations. Figure 8 shows examples of the 25 × 2 faces generated from the training set for two people. From the automatically generated faces, it is clear that some of them have degradations, like getting two noses, or they are highly blurred. Some images represent the integration of two or three facial poses. Figure 9 shows the ten original faces included in the ORL database at the top, and the resulting 25 faces after training the LVQ classifier. The 25 faces are generated by plotting the weight vectors of the hidden layer neurons, which represent the ten faces. Close investigation of the 25 faces indicates that some of the faces seem to have got glasses, although the original faces had no glasses. This indicates that the LVQ generalises very well to reflect all the possible sources of variability.

5.1.4. Enhancing the Generalisation and Rejection of the LVQ Classifier. Image-based synthesis of facial images has useful application in enhancing the performance of face classifiers. The faces, which normally interfere with each other, could be used to form a new set of faces, which are used as training patterns in the reject class. Faces which are normally falsely accepted by being assigned to the wrong class could be assigned to approximately similar faces in the reject class, as it includes greater similarity to them than the other faces. To enhance the performance of face recognisers, synthesised faces are used. Consider an M+1 class classifier C with a reject class. The patterns of the two classes C_i and C_j interfere with each other. This means that the two classes are nearer to each other than to the reject class. To avoid such confusion, merging pairs of the confused classes using an LVQ classifier generates new faces. The resulting faces have a certain degree of similarity to both of the original faces. Including these newly generated faces in the reject class attracts the patterns which are normally falsely accepted as members of other classes into the reject class. This process is performed after a preliminary test phase on the test set. The confused classes are then specified to generate common faces, which are then added to the reject class. This process results in enhancing the system's rejection performance and reducing the false acceptance rate.

5.2. Experiments on the RBF Classifier

The effects of spread constant and image size on the performance of the RBF classifier are studied. The only para-

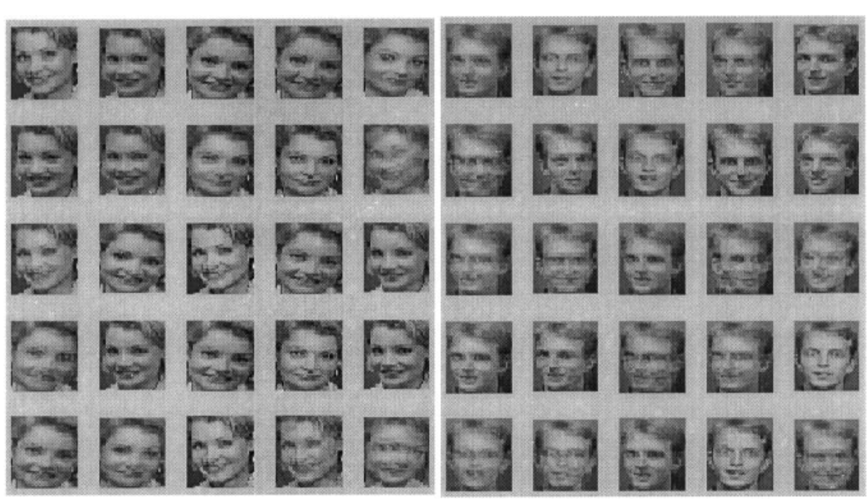

Fig. 8. Artificially generated faces using a LVQ network.

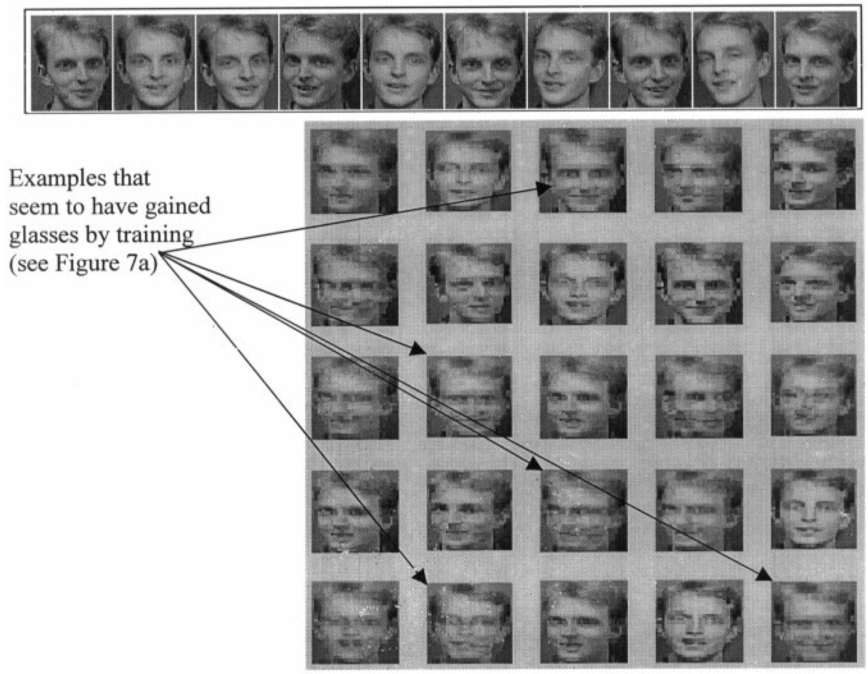

Fig. 9. Original faces in the ORL database and the 25 weight vectors generated after training the LVQ classifier.

Table 3. RBF classifier performance (% PCR) for different spread constant (S)

S	1	4	5	5.5	6	7	8	20	30	40	60	100
PCR	0	96	98	97.5	97.5	96.5	95	93	90.5	89.5	88.5	89.5

meter that greatly affects the generalisation performance of the radial basis function classifier is the spread constant. Many experiments were performed to conduct the most suitable constant. Table 3 shows the classifier performance for different values of the spread constant (S).

Increased image size imposes a heavy computational load, and needs more training time. Table 4 indicates that the spread constant of the radial basis neuron is affected by the size of the training patterns.

To demonstrate the generalisation of the RBF classifier, the faces generated as the outputs of some RBF neurons are shown in Fig. 10. These faces indicate the good generalisation capability of RBFs.

5.3. Experiments of the Combined Classifier

The best performing classifiers are selected for building a combined classifier (see Fig. 5). A learning vector quantis-

Table 4. RBF classifier performance for different sizes

Image size	32×32	64×64
Spread constant	5	5
PCR	98	62.5

Fig. 10. Outputs of some trained radial basis neurons.

Table 5. Performance of different classifiers

Set	PCR		
	LVQ classifier	RBF classifier	Combined classsifier LVQ+RBF+FEC
Training	100	100	100
Test	99.0	98	99.5%

ation network with 1200 hidden neurons resulted in 99% correct classification. A RBF network trained on the faces of the confused classes resulted in only one misclassification. Combining the results of both classifiers, the system performance is improved to a correct recognition rate of 99.5%, with a rejection rate of 0.5% based on the following set of rules, which is used to set the target classes of the front end classifier:

1. If the pattern is rejected by the LVQ classifier then reject it (note that rejection performance of LVQ is the best).
2. If the pattern is assigned to a class by the LVQ and rejected by the special RBF classifier, then the class is that class decided by the LVQ classifier.
3. If classified by both the LVQ and the RBF classifiers to the same class, then accept that class.
4. If both the LVQ classifier and the RBF classifier disagree about the class assigned to a new pattern, then accept the class to be that of the RBF classifier.

These rules resulted in 99.5% correct recognition rate at a rejection rate of 0.5%. Table 5 summarises the results of the different classifiers.

6. DISCUSSION AND COMPARISON WITH PREVIOUS TECHNIQUES

Extensive experiments were executed on a 200 MHz Intel processor with MMX running in a Windows environment. The average recognition time per face is 0.463 seconds (Table 6). The disadvantage of the RBF approach is the linear dependency of computational complexity on the number of radial basis neurons (number of training patterns). The advantage of RBF approach is that the addition of new training patterns could be easily done without retraining of the network.

Table 6. Time requirements of the different classifiers

Processing phase	Computation Time, Seconds		
	LVQ classifier	RBF classifier	Combined classifier
Size normalisation	0.026	0.026	0.026
Recognition	0.373	0.256	0.437
Total	0.399	0.282	0.463

Table 7 also indicates that using several views of the faces as a training set increases the general recognition performance of the classifier, as well as its ability to discriminate between learned faces presented from a new view angle and new faces. Experiments have shown that at least five training faces per individual were necessary for acceptable performance.

7. SUMMARY AND CONCLUSIONS

Invariant face recognition is a challenging task in computer vision. The selection of stable and representative sets of features that efficiently discriminate between faces in a huge database is the major problem. Variability of facial pose, expressions and lighting conditions render it very difficult to extract such a representative and stable set of features using geometrical methods. On the other hand, neural networks represent the most suitable way to automatically extract such a set. A neural network plays a two-fold rule: feature extractor and classifier at the same time. This paper presented the design and implementation issues of a practical system for invariant face recognition. The system showed an outstanding performance compared to state-of-the-art systems, which were tested on the ORL database.

In this paper, a category-based combined classifier has been proposed to improve the generalisation capability, and hence the system performance. Combining different classifiers can improve the overall system performance, even if the individual classifiers agree on the wrong classification decision. In this research, a new idea that is based on combining two different specialised classifiers is proposed. The proposed classifier uses the generalisation capabilities of both the LVQ and RBF classifiers to overcome the limitations of the used individual classifiers. While, the LVQ classifier is trained on the set of distinctive faces, the RBF classifier is trained on the set of familiar faces. Training the RBF classifiers on a rejection class, which includes synthesised faces from the members of the familiar faces class, results in enhancing the rejection performance of the system. The inner workings of the neural networks are visualised in order to get insights about the learning process.

Both the LVQ and the RBF neural networks proved successful in generalisation for invariant face recognition. The RBF network is faster than the LVQ network, but the LVQ performance is slightly better (1.0%) than that of the RBF network. Although the individual classifiers perform better than other neural networks reported in the overview, their combination results in a better performance.

The following conclusions could be drawn:

- Combined classifiers that are based on neural networks offer significant improvements over the component classifiers.
- The selection of the number and pose of the training

Table 7. Error rates for different numbers of training faces per subject

Training set/individual	1	2	3	4	5	6	7	8
Convolutional network + SOM [15]	30.0	17.0	11.8	7.1	3.8			
VFR-Model [20]					7.5	4.4	3.4	0.0
Eigenfaces – average per class [15]	38.6	28.8	28.9	27.1	26			
Eigenfaces – one per image [15]	38.6	20.9	18.2	15.4	10.5			
PCA+CNN [15]	34.2	17.2	13.2	12.1	7.5			
HMM [13]					5			
LVQ (this paper)				5	1	0.0	0.0	0.0
RBF (this paper)				6	2	1	1	0.0
LVQ+RBF+FEC (this paper)					0.5	0.0	0.0	0.0

faces per individual is crucial to the recognition process. At least five training faces, which encompass all the possible varieties, are necessary to achieve acceptable system performance of 99.5% for the ORL database.

- Specialised classifier networks resulted in enhancing the system performance. The experimental results reported in this research are consistent with those reported by David [11]. David [11] suggests that the composite classifiers which result from combining a small number of component classifiers, where each component stores a small number of prototypical instances, is more accurate than a classifier that stores all training instances as prototypes. While in David [11] algorithms that rely primarily on random sampling are used to select a small number of prototypes, we used a preliminary training stage for the LVQ classifier for separating the distinctive faces from the familiar ones. Each set of faces is used subsequently in training one of the best performing component classifiers.
- The combined classifier implemented in this research may be applied to similar recognition tasks.
- The face generation mechanism presented could be applied to generate faces for the reject class, to enhance the rejection performance in face recognition systems.
- The ORL database is proved to encompass much variability than the most widely used FERRET database. The results reported by other researchers [15,17] indicate that the approaches, which are highly efficient for recognising the faces of the FERRET database, fail to cope with the conditions of the ORL database.

Our future work will focus on extending our system to cope with large databases, and developing an automatic approach for optimal selection of the combined classifier components.

Acknowledgements

This work was supported by research grant SP053 received from the Research Administration in Kuwait University. The author would like to thank the Olivetti Research Laboratory and Ferdinando, Samaria for compiling and maintaining the (ORL) database.

References

1. Hoogenboom R. Human Face Detection and Recognition. Masters Thesis, University of Leiden, 1996
2. Sung KK, Poggio T. Example-based learning for view-based human face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 1997; 20(1):39–51
3. Daugman J. Face and gesture recognition: Overview. IEEE Transactions on Pattern Analysis and Machine Intelligence 1997; 19(7):675–676
4. Rowley HA, Baluja S, Kanade T. Neural network-based face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 1998; 20(1):23–38
5. Chellappa R, Wilson CL, Sirohey S. Human and machine recognition of faces: A survey. Proceedings of the IEEE 1995; 83(5):705–740
6. Valentin D, Abdi H, Toole AJO, Cottrell W. Connectionist models of face processing: a survey. Pattern Recognition 1994; 27:1209–1230
7. Samal A, Iyangar P. Automatic recognition and analysis of human faces and facial expressions: A survey. Pattern Recognition 1992; 25(1):65–77
8. Ho TK, Hull JJ, Srihari SN. Decision combination in multiclassifier systems. IEEE Transactions on Pattern Analysis and Machine Intelligence 1996; (I)6:66–75
9. Srinivas G, Wechsler H. Face recognition using hybrid classifiers. Pattern Recognition 1997; 30(4):539–553
10. Zhang M, Flucher J. Face recognition using artificial neural network Group-based Adaptive Tolerance (GAT) trees. IEEE Transactions on Neural Networks 1996; 7(3):555–567
11. David BS. Prototype Selection for Composite Nearest Neighbor Classifiers. PhD Dissertation, University of Masachusetts Amherst, 1997
12. Chuanyi JI, Sheng Ma. Combinations of weak classifiers. IEEE Transactions on Neural Networks, 1997; (8)1:32–42
13. Samaria AS. Face Recognition Using Hidden Markov Models. PhD Dissertation, Trinity College, University of Cambridge, UK, 1994
14. Turk M, Pentland A. Eigenfaces for recognition. Journal of Cognitive Neuroscience 1991; (3)1:71–86
15. Lawrence S, Lee GC, Tsoi AC, Back A. Andrew Face Recognition: A convolutional neural network approach. IEEE Transactions on Neural Networks 1997; (8)1:98–113
16. Belhumeur PN, Hespanha JP, Kreigman DJ. Eigenfaces ve. Fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 1997, 18(7):71–720

17. Lin SH, Kung SY, Lin LJ. Face recognition/detection by probabilistic decision-based neural network. IEEE Transactions on Neural Networks 1997; (8)1:114–132
18. Feitosa RQ, Thomaz CE, Veigo A. Comparing the performance of the Discriminant Analysis and RBF Neural Network for Face Recognition. Proceedings 5th International Conference on Information Systems Analysis and Synthesis, 1999; 6:50–56
19. Hagen S. Face Recognition – a novel technique. Master Thesis, University of Heidelberg, Germany, 1995
20. Ben-Arie J, Nandy D. A volumetric/iconic frequency domain representation for objects with application for pose invariant face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 1998; (20)5:449–457
21. Kin-Man L, Hong Y. An analytic-to-holistic approach for face recognition based on a single frontal view. IEEE Transactions on Pattern Analysis and Machine Intelligence 1998; (20)7
22. Li SZ, Lu J. Face recognition using the nearest feature line method. IEEE Transactions on Neural Networks 1999; (10)2:439–443
23. Sutherland K, Renshaw D, Denyer BP. A novel automatic face recognition algorithm employing vector quantization. IEEE CDs
24. Wiskott L, Fellous JM, Krueger N, von der Malsburg C. Face recognition by elastic bunch graph matching. IEEE Transactions on Pattern Analysis and Machine Intelligence 1997; 19(7):665–779
25. We S. Face Recognition. (http://www.cim.mcgill.ca/~wsun/sa/project/report.html])
26. Biship CA. Neural Networks for Pattern Recognition. 1st Edition. Clarendon Press, Oxford, 1995
27. Valentin A, Abdi H. Can a linear autoassociator recognize faces from new orientations. Journal of Optical Society of America 1996
28. Gross M. Visual Computing. Springer-Verlag, 1995; 289
29. Kohonen T. Self-Organizing Maps. Springer-Verlag, 1995
30. Bandya AS, Macy RB. Pattern Recognition with Neural Networks in C++. CRC Press, 1996
31. Hush DR, Horne BG. Progress in supervised neural networks. IEEE Signal Processing Magazine 1993; 5(1):8–39
32. Chen S, Cowan CFN, Grant PM. Orthogonal least squares learning algorithm for radial bias function networks. IEEE Transactions on Neural Networks 1991; 2(2):302–309
33. Kittler J, Hatef M, Duin RPW, Matas J. On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Ingelligence 1998; (20)3:226–239
34. Bruzzone L, Pierto DF. A technique for the selection of kernel-function parameters in RBF neural networks for classification of remote-sensing images. IEEE Transactions on Geoscience and Remote sensing 1999; 37(2)

A.S. Tolba received his BSc (with honours) and MSc degrees from the University of Almansoura, Egypt, in 1978 and 1981, respectively. He received his PhD from the University of Wuppertal, Germany, in 1988. Since 1994 he has been an Associate Professor of Computer Engineering at the University of Suez-Canal, Egypt. He is currently on sabbatical leave to the Department of Physics at the University of Kuwait. He has done research in compuer vision, biometrics identification, autonomous vehicles, neural networks and gesture recognition. He has published more than 40 papers in these areas. He is co-author of two edited books: *Intelligent Robotic Systems* (Marcel Dekker, New York, 1991) and *Laser Technology and its Applications* (publication of ISESCO, 1997). His most recent research focus is face recognition and gesture recognition. Dr Tolba is a member of the IEEE and AMSE societies. He is currently a member of the editorial board of *Modeling, Measurement and Control*.

Abdulnasser Abu-Rezq received his BSc degree in applied physics and digital electronics from Kuwait University in 1980. He received a PhD degree from Kent University at Canterbury, UK in 1985. He is currently with the Department of Physics at the University of Kuwait State. His research interests include digital electronics, image processing, pattern recognition and neural networks. Dr Abu-Rezq is a member of the IEEE Society.

Correspondence and offprint requests to: A.S. Tolba, Physics Department, Kuwait University, PO Box 5969, Safat 13060, Kuwait. Email: Tolba@kuc01.kuniv-.edu.kw