Invariant Recognition Of Human Faces

GULZAR A. KHUWAJA
Physics Department
Kuwait University
P. O. Box 5969, Safat 13060
KUWAIT

Abstract: - Visual communication plays an important role in human communication and interaction. In order to interact socially, we must be able to process faces in a variety of ways. In this paper, an algorithm for invariant recognition of human faces based on LVQ neural network is presented. The proposed system is shown to exhibit robustness in achieving better classification results with both good generalization performance and a fast training time on a variety of test problems using hundreds of faces. Sources of variability include facial expression, gender, individual appearance, tilt, lighting conditions, and occluding objects (hair, spectacles, etc).

Key-Words: - Pattern Recognition, Computer Vision, Adaptive Classification, Invariant Face Recognition.

1 Introduction
Invariant face classification is a challenging task, especially in the absence of highly controlled environments and recognition constraints. Although the process for recognizing faces has received considerable attention in the course of computer vision research [1,2], the present stage is far from the goal of humanlike capability especially from the point of robustness to wide-range lighting. The daily shadow, reflection and darkness have great influence on the accuracy of facial recognition through the inevitable change of greylevel. Pattern variability often causes poor pattern recognition accuracy. Earlier publications by the authors reported findings on pattern recognition based on non-adaptive feature fusion [3] and decision fusion [4]. The reported investigations showed that the use of multiple information sources, exhibiting redundancy or complementarity itself does not result in robust pattern recognition. The studies showed that mismatched recognition and fusion-training conditions lead to poorer recognition accuracy than matched conditions. This suggests that robust pattern recognition may require adaptation [5].

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 and various statistical methods including cross-validation and model selection. Neural networks are parallel computational models, with varying degrees of complexity, comprised of densely interconnected adaptive processing units. These networks are fine-grained parallel implementations of nonlinear static or dynamic systems. A very important feature of these networks is their adaptive nature of learning by example in solving problems. This feature makes such computational models very appealing for a wide variety of application.

2 LVQ Classifier
LVQ (learning vector quantization) is a supervised classifier that was first studied by Kohonen [6]. 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. 1) 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 \( \mathbf{X}_j \), a winner neuron \( W_i \) is chosen to adjust its weight vector:

\[
\| \mathbf{X}_j - W_i \| \leq \| \mathbf{X}_j - W_k \|, \quad \text{for all } k \neq i
\]

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

\[
W_i(t+1) = W_i(t) + \alpha (\mathbf{X}_j - W_i(t))
\]

if \( \mathbf{X}_j \) and \( W_i \) belong to the same class

\[
W_i(t+1) = W_i(t) - \alpha (\mathbf{X}_j - W_i(t))
\]

if \( \mathbf{X}_j \) 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 [6]. Otherwise, do not change the weights. The training algorithm is stopped after reaching a pre-specified error limit. 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 as the winner. This in turn fires one output neuron, signifying a particular class.

![Architecture of the LVQ Classifier](image)

### 3 Databases and Preprocessing

#### 3.1 KU Database

The KU database [7] was developed at the Kuwait University, Kuwait. This database contains 240 images posed by 48 subjects. Each subject took 5 pictures of herself or himself. Each face image is of resolution 128 x 128 pixels and 8-bit greylevel bmp format (Windows Bitmap). There are variations in facial expression (open/closed eyes, smiling/non-smiling), facial details (glasses/no glasses), and rotation of up to 60.

#### 3.2 ORL Database

The ORL database [8] was developed at the Olivetti Research Laboratory, Cambridge. The data consists of 400 images acquired from 40 persons, some of which were taken at different times for some of the persons. 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 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 gray scale with a resolution of 92 x 112 pixels. The images are size normalized.

#### 3.3 PICS Database

The Psychological Image Collection at Stirling (PICS) database [9] was developed at the Psychology Department, University of Stirling, UK. It consists of many data sets. The data set selected for this study is ntt-faces-originals. The database comprises on 490 images of 70 subjects. Each face is posed in four expressions with a frontal view, two expressions with a 3/4 view, and one frontal view wearing a bathing cap. The images are available in 8-bits greylevel gif format of variable sizes. The cropped face images are used for the recognition system.

### 4 Algorithm

The most challenging step in the design of a pattern recognition system is the selection of a suitable base model that constitutes its building blocks. The next step is the features selection and extraction method. The selection of the face image only makes the proposed architecture feasible in real-time application domains. The network classifier for face recognition is trained on the random training sets. The training samples are individually sized specifically to get the overall performance scenario of the network architecture.

#### 4.1 LVQ Models

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 (HN) should be suitable to capture the knowledge of the problem domain. For example, training a neural
network to recognize faces which belong to number of classes (NC), 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.

Visualizing the learned pattern of the hidden layer neurons, it is found that there are neurons with completely blurred patterns, blind neurons [10], as these neurons did not see the faces which are clamped to the neurons of the input layer. Eliminating the blind neurons enhances the classifier performance. The algorithm based on efficient LVQ model parameters is as follows:

1- Select the network parameters:
   - Input layer size = Image size (32 x 32 = 1024 neurons).
   - Training set size = S (10 or 20 subjects) * X (2 or 3 or 4 faces).
   - Number of classes (NC) = S (Number of subjects).
   - Hidden layer neurons (HN) = S * Round(X/2).
   - Learning rate (α) = 0.1.

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 or 1500 or 2500.

2- Initialize an LVQ classifier: Initialization of the weight matrix for competitive layer w1 and linear layer w2.

3- Start training of an LVQ classifier based on selected efficient model parameters.

4- Test the trained classifier on both training and test sets and compute pcctr and pccts (percentage of correct classification for training and test sets respectively).

5- Exit.

4.2 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.

Face selection is an important problem when designing a pattern recognition system that is concerned with which attributes are most relevant for decision-making. Features that 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.

5 Empirical Results
Many experiments are performed to explore the possibility of the best parameters selection for invariant face recognition of an adaptive neural network. Tables 1-3 show various results of the network architecture trained for the recognition of human faces for KU, ORL, and PICS databases [7-9] respectively. As shown in Table 1, the recognition rate of the network architecture is 100% for a random test set of 20 subjects with 60 faces, i.e., three faces per subject for KU database [7]. The system is trained with a training set of 40 images of same subjects. The efficiency of the proposed architecture is evaluated on both the time and the space scale. By setting the number of hidden layer units (HN) equal to the S* Round(X/2) has reduced the network memory requirements for the internal representation of target faces. In addition, it enhances the processing speed, both training time of the network and recognition time of the face image are reduced. The recognition time (cpu) per face image on the average is about 0.04 seconds, which makes the proposed architecture feasible for large data training and test samples in real-time application domains.

<table>
<thead>
<tr>
<th>S</th>
<th>EP</th>
<th>I</th>
<th>Tr</th>
<th>Ts</th>
<th>NH</th>
<th>peak</th>
<th>10 run</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1000</td>
<td>50</td>
<td>20</td>
<td>30</td>
<td>10</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>20</td>
<td>2500</td>
<td>100</td>
<td>40</td>
<td>60</td>
<td>20</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

S = subjects, EP = epochs, I = images (total), Tr = training set, Ts = test set, NH = number of hidden neurons, IS = image size, pcctr = percentage of correct classification of training set, pccts = percentage of correct classification of test set
degraded. However, this infers that the architecture is quite capable of recognizing more target faces efficiently and is independent of the source and size of training data. We have divided databases into different size random training and test sets to get the clear picture of the proposed network architecture's efficiency and performance.

Table 2. The percentage of correct classification of test sets of face recognition for ORL database with parameters: $\alpha = 0.1$, IS = 32x32 pixels.

<table>
<thead>
<tr>
<th>S</th>
<th>EP</th>
<th>I</th>
<th>Tr</th>
<th>Ts</th>
<th>NH</th>
<th>peak pctcs</th>
<th>10 run pctcs</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1000</td>
<td>100</td>
<td>40</td>
<td>60</td>
<td>20</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>20</td>
<td>2500</td>
<td>200</td>
<td>80</td>
<td>120</td>
<td>40</td>
<td>99.16</td>
<td>98.51</td>
</tr>
</tbody>
</table>

Table 3 provides similar results for PICS database [9]. The recognition rate of the network architecture is 98.1% on the average of 10 run for a random test set of 10 subjects with 30 faces. The recognition time (cpu) per face image on the average is about 0.02 seconds. Redundant hidden layer neurons not only decrease the efficiency of the network system but also the performance as many units not evolved properly during the training phase create confusion in the decision making process.

Table 3. The percentage of correct classification of test sets of face recognition for PICS database with parameters: $\alpha = 0.1$, IS = 32x32 pixels.

<table>
<thead>
<tr>
<th>S</th>
<th>EP</th>
<th>I</th>
<th>Tr</th>
<th>Ts</th>
<th>NH</th>
<th>peak pctcs</th>
<th>10 run pctcs</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1000</td>
<td>60</td>
<td>30</td>
<td>30</td>
<td>20</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>20</td>
<td>2500</td>
<td>120</td>
<td>60</td>
<td>60</td>
<td>40</td>
<td>98.33</td>
<td>98.10</td>
</tr>
</tbody>
</table>

The performance of a supervised learning system is also characterized by its generalization error, 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 [11]. That is, if a neural network is too large, may over-fit a particular training set and fail to maintain good generalization error. A small neural network, however, may be sufficient to approximate an optimal solution.

The generalization superiority of the neural net can be attributed to the bounded and smooth nature of the hidden-unit responses. Once particular approximation function or network architecture is decided on, generalization can be improved if the number of free parameters in the net is optimized [12]. Thus, eliminating the redundant hidden layer neurons of the network architecture not only makes the system more efficient but also increases the system performance.

6 Summary

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 is that they can extract the most discriminative and representative set of features.

We have presented a learning vector quantization neural network architecture based on varying parameters and eliminating redundant hidden layer units or blind neurons [10] that learns the correlation of patterns and recognizes human faces. The network classifier is trained on the random training samples to perform recognition task on the input face image. Empirical results yield an accuracy rate of 100% for a random test set of 60 face images of 20 subjects on the network that is trained with another set of 40 faces of same subjects. The recognition time is about 0.04 seconds per face image on an IBM PC. The empirical results of the proposed network architecture trained with one set of faces and tested with a new set of facial varieties which were not included in the training sample are given in Tables 1-3.

References: