Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 468693, 7 pages
doi:10.1155/2008/468693

Research Article

Comparing Robustness of Pairwise and Multiclass Neural-Network Systems for Face Recognition

J. Uglov, L. Jakaite, V. Schetinin, and C. Maple

Computing and Information System Department, University of Bedfordshire, Luton LU1 3JU, UK

Correspondence should be addressed to V. Schetinin, vitaly.schetinin@beds.ac.uk

Received 16 June 2007; Revised 28 August 2007; Accepted 19 November 2007

Recommended by Konstantinos N. Plataniotis

Noise, corruptions, and variations in face images can seriously hurt the performance of face-recognition systems. To make these systems robust to noise and corruptions in image data, multiclass neural networks capable of learning from noisy data have been suggested. However on large face datasets such systems cannot provide the robustness at a high level. In this paper, we explore a pairwise neural-network system as an alternative approach to improve the robustness of face recognition. In our experiments, the pairwise recognition system is shown to outperform the multiclass-recognition system in terms of the predictive accuracy on the test face images.

1. INTRODUCTION

The performance of face-recognition systems is achieved at a high level when these systems are robust to noise, corruptions, and variations in face images [1]. To make face recognition systems robust, multiclass artificial neural networks (ANNs) capable of learning from noisy data have been suggested [1, 2]. However, on large face image datasets, containing many images per class (subject) or large number of classes, such neural-network systems cannot provide the performance at a high level. This happens because boundaries between classes become complex and a recognition system can fail to solve a problem; see [1–3].

To overcome such problems, pairwise classification systems have been proposed; see, for example, [4]. Pairwise classification system transforms a multiclass problem into a set of binary classification problems for which class boundaries become much simpler than those for a multiclass system. Beside that, the density of training samples for a pairwise classifier becomes lower than that for a multiclass system, making a training task even simpler. As a result, classifiers in a pairwise system can learn to divide pairs of classes most efficiently.

The outcomes of pairwise classifiers, being treated as class membership probabilities, can be combined into the final class posteriori probabilities as proposed in [4]. This proposed method aims to approximate the desired posteriori probabilities for each input although such an approximation requires additional computations. Alternatively, we can treat the outcomes of pairwise classifiers as class membership values (not as probabilities) and then combine them to make decisions by using the winner-take-all strategy. We found that this strategy can be efficiently implemented within a neural network paradigm in the competitive layer as described in [5].

However, the efficiency of such pairwise neural-network schemes has not been yet explored sufficiently in face recognition applications. For this reason in this paper we are aiming to explore the ability of pairwise neural-network systems to improve the robustness of face recognition systems. The exploration of this issue is very important in practice, and that is the novelty of this research. In our experiments, the pairwise neural networks are shown to outperform the multiclass neural-network systems in terms of the predictive accuracy on the real face image datasets.

Further in Section 2, we briefly describe a face image representation technique and then illustrate problems caused by noise and variations in image data. Then in Section 3 we introduce a pairwise neural-network system proposed to enhance the robustness of face recognition system. In Section 4 we describe our experiments, and finally in Section 5 we conclude the paper.

2. FACE IMAGE REPRESENTATION AND NOISE PROBLEMS

Image data are processed efficiently when they are represented as low-dimensional vectors. Principal component analysis (PCA), allowing data to be represented in a low-dimensional space of principal components, is a common technique for image representation in face recognition systems; see, for example, [1–3]. Resultant principal components make different contribution to the classification problem.

The first two principal components, which make the most important contribution to face recognition, can be used to visualise the scatter of patterns of different classes (faces). Particularly, the use of such visualisation allows us to observe how noise can corrupt the boundaries of classes. For instance, Figure 1 shows two examples of data samples representing four classes whose centres of gravity are visually distinct. The left-side plot depicts the samples taken from the original data while the right-side plot depicts the same samples mixed with noise drawn from a Gaussian density function with zero mean and the standard deviation alpha = 0.5.

From the above plot, we can observe that the noise corrupts the boundaries of the classes, affecting the performance of a face recognition system. It is also interesting to note that the boundaries between pairs of the classes do not change much. This observation inspires us to exploit a pairwise-classification scheme to implement a neural network-based face recognition system which would be robust to noise in image data.

3. A PAIRWISE NEURAL-NETWORK SYSTEM FOR FACE RECOGNITION

The idea behind the pairwise classification is to use two-class ANNs learning to classify all possible pairs of classes. Consequently, for C classes a pairwise system should include $C^*(C-1)/2$ ANNs trained to solve two-class problems. For instance, given $C = 3$ classes Ω_1, Ω_2, and Ω_3 depicted in Figure 2, we can setup three two-class ANNs as illustrated in this figure. The lines $f_{i/j}$ are the separating functions learnt by the ANNs to separate class i from class j. We can assume that functions $f_{i/j}$ give the positive values for inputs belonging to classes i and the negative values for the classes j.

Now we can combine functions $f_{1/2}$, $f_{1/3}$, and $f_{2/3}$ to build up the new separating functions g_1, g_2, and g_3. The first function g_1 combines the outputs of functions$f_{1/2}$ and $f_{1/3}$ so that $g_1 = f_{1/2} + f_{1/3}$. These functions are taken with weights of 1.0 because both $f_{1/2}$ and $f_{1/3}$ give the positive output values for data samples of class Ω_1. Likewise, the second and third separating functions g_2 and g_3 are described as follows:

$$g_2 = f_{2/3} - f_{1/2}, \qquad g_3 = -f_{1/3} - f_{2/3}. \tag{1}$$

In practice, each of the separating functions $g_1, \ldots, g_c$ can be implemented as a two-layer feed-forward ANN with a given number of hidden neurons fully connected to the input nodes. Then we can introduce C output neurons summing all outputs of the ANNs to make a final decision. For instance, the pairwise neural-network system depicted in Figure 3 consists of three ANNs learning to approximate functions $f_{1/2}$, $f_{1/3}$, and $f_{2/3}$. The three output neurons g_1, g_2, and g_3 are connected to these networks with weights equal to (+1, +1), (−1, +1), and (−1,−1), respectively.

In general, a pairwise neural-network system consists of $C(C-1)/2$ ANN classifiers, represented by functions $f_{1/2}, \ldots, f_{i/j}, \ldots, f_{C-1/C}$, and C output neurons $g_1, \ldots, g_c$, where $i < j = 2, \ldots, C$. We can see that the weights of output neurons g_i connected to the classifiers $f_{i/k}$ and $f_{k/i}$ should be equal to +1 and −1, respectively.

Next, we describe the experiments which are carried out to evaluate the performance of this technique on synthetic and real face images datasets. The performances of the pairwise-recognition systems are compared with those of the multiclass neural networks.

4. EXPERIMENTS

In this section, we describe our experiments with synthetic and real face image datasets, aiming to examine the proposed pairwise and multiclass neural-network systems. The examination of these systems is carried out within 5-fold cross-validation.

4.1. *Implementation of recognition systems*

In our experiments, both pairwise and standard multiclass neural networks were implemented in Matlab, using neural networks Toolbox. The pairwise classifiers and the multiclass networks include hidden and output layers. For the pairwise classifiers, the best performance was achieved with two hidden neurons, while for the multiclass networks the numbers of hidden neurons were dependent on problems and ranged between 25 and 200. The best performance for pairwise classifiers was obtained with a tangential sigmoid activation function (tansig), while for multiclass networks with a linear activation function (purelin). Both types of the networks were trained by error back-propagation method.

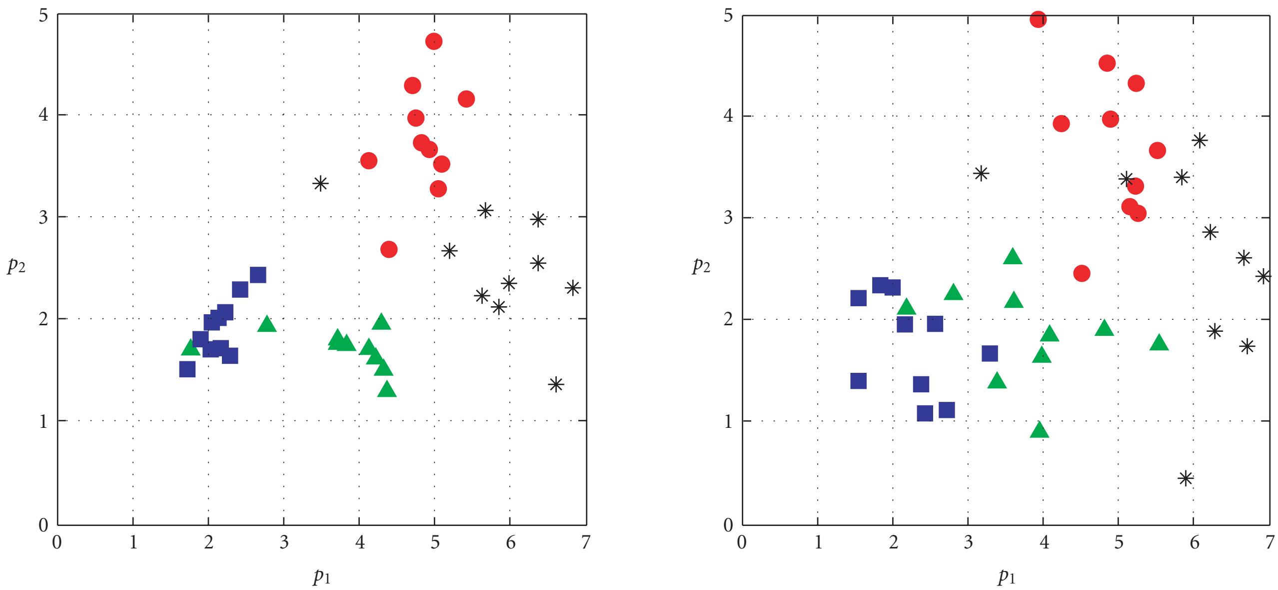
4.2. *Face image datasets*

All the face images used in our experiments are processed to be in a grey scale ranging between 0 and 255. Because of large dimensionalities of these data, we used only the first 100 principal components retrieved with function “princomp”.

The face image datasets Cambridge ORL [6], Yale extended B [7], and Faces94 [8], which were used in our experiments, were partially cropped and resized in order to satisfy the conditions of using function “princomp”. Image sizes for the ORL, Yale extended B, and Faces94 were 64×64, 32×32, and 45×50 pixels, respectively. For these face image sets, the number of classes and number of samples per subject were 40 and 10, 38 and 60, and 150 and 20, respectively.

4.3. *Impact of data density in case of synthetic data*

These experiments aim to compare the robustness of the proposed and multiclass neural networks to the density of



(a) (b)

FIGURE 1: An example of scattering the samples drawn from the four classes for alpha = 0 (a) and alpha = 0.5 (b) in a plane of the first two principal components p_1 and p_2.

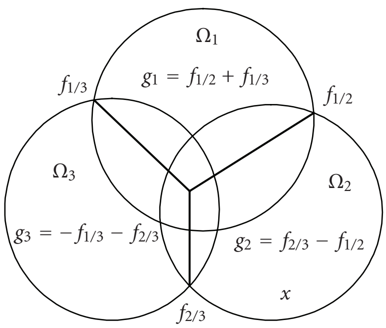


FIGURE 2: Splitting functions $f_{1/2}, f_{1/3}$, and $f_{2/3}$ dividing the following pairs of classes: Ω_1 versus Ω_2, Ω_1 versus Ω_3, and Ω_2 versus Ω_3.

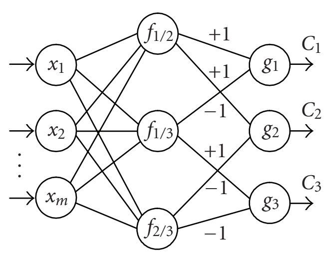


FIGURE 3: An example of pairwise neural-network system for $C = 3$ classes.

synthetic data. The synthetic data were generated for four classes which were linearly separable in a space of two variables, p_1 and p_2 that allowed us to visualise the boundaries between the classes. Each of these variables ranges between 0 and 1.

The class boundaries are given by the following lines:

$$y = p_1 + 0.5, \qquad y = p_2 + 0.5. \tag{2}$$

The number of data samples in each class was given between 10 and 200, making the data density different. Clearly, when the density is higher, the data points are closer to each other, and the classification problem becomes more difficult. Figure 4 shows two cases of the data densities with 10 and 200 samples per class.

From this figure, we see that when the density is high the data samples may be very close to each other, making the classification problem difficult. Hence, when the data density is high or the number of classes is large, pairwise classifiers learnt from data samples of two classes can outperform multiclass systems learnt from all the data samples. This happens because the boundaries between pairs of classes become simpler than the boundaries between all the classes.

The robustness of the proposed pairwise and multiclass systems is evaluated in terms of the predictive accuracy on data samples uniformly distributed within (0, 1). The classes $C_1, \dots, C_4$ are formed as follows:

$$\begin{gathered} C_1 : p_1 \in [0, 0.5], \quad p_2 \in [0, 0.5]; \quad C_2 : p_1 \in [0, 0.5], \\ p_2 \in [0.5, 1.0], \quad C_3 : p_1 \in [0.5, 1.0], \quad p_2 \in [0.5, 1.0]; \\ C_4 : p_1 \in [0.5, 1.0], \quad p_2 \in [0, 0.5]. \end{gathered} \tag{3}$$

In theory, multiclass neural networks with two hidden and four output neurons are capable of solving this classification problem. However, practically the performance of a multiclass neural network is dependent on the initial weights as well as on the density of data samples.

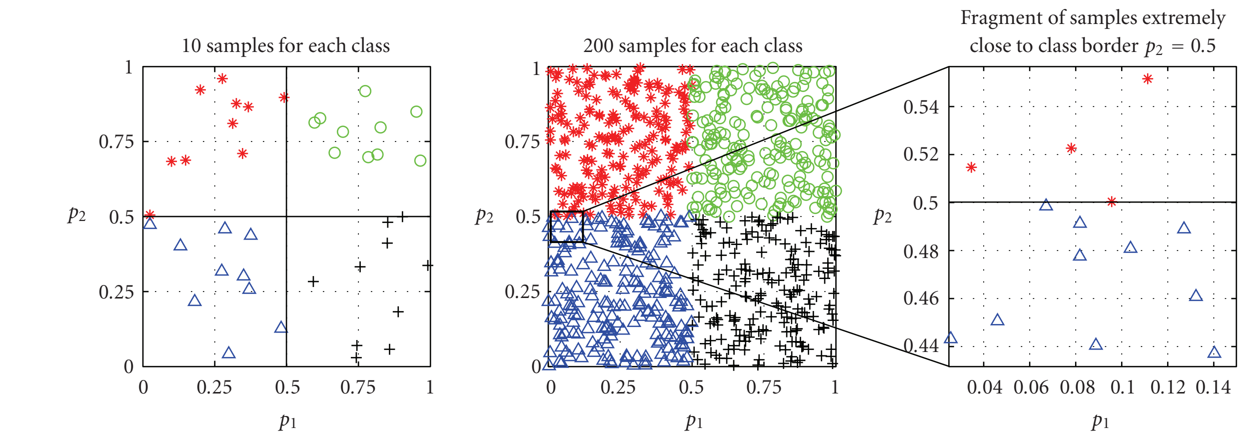


Figure 4: High density of data samples makes the classification problem difficult. The zoomed fragment shows how close are the data samples to each other.

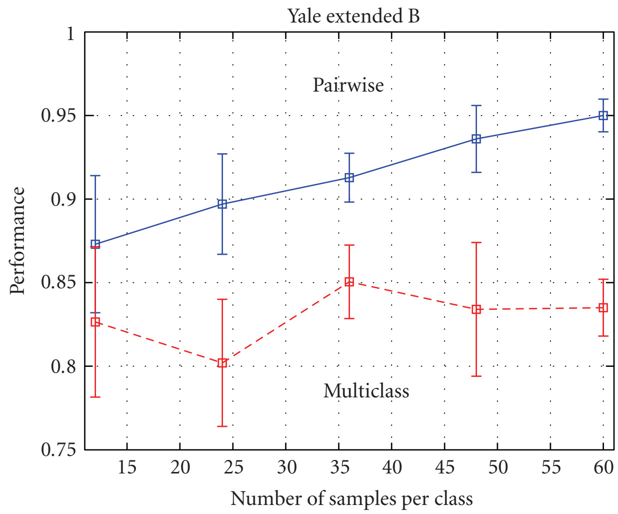


Figure 5: Performances of the pairwise and multiclass recognition systems versus the numbers of samples per subject. Solid lines and bars are the mean and 2σ intervals, respectively.

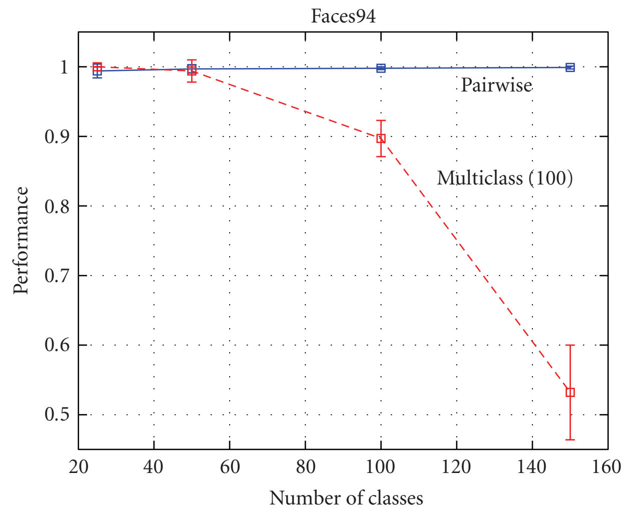


Figure 6: Performance of the pairwise and multiclass-recognition systems over the number of classes. Solid lines and bars are the mean and 2σ intervals, respectively.

In our experiments, the numbers of data samples per class were given between 50 and 200. Table 1 shows the performances of the pairwise and multiclass systems for these data.

From this table we can see that the proposed pairwise system outperforms the multiclass system on 16% and 20% when the numbers of samples are 50 and 200, respectively.

4.4. Impact of data density in case of Yale data

The Yale extended B data contain 60 samples per subject that gives us an opportunity to examine the robustness of the face recognition systems to the data density. In these experiments, we compare the performances of both recognition systems trained on the datasets containing different number of samples per subject. The numbers of these samples are given 12, 24, 36, 48, and 60 per subject. Figure 5 shows the performance of the proposed pairwise and multiclass systems over the number of samples per subject.

From this figure, we can see that the proposed pairwise-recognition system significantly outperforms the multiclass system in terms of the predictive accuracy on the test data. For instance, for 24 samples a gain in the accuracy is equal to 9.5%. When the number of samples is 60, the gain becomes 11.5%.

4.5. Impact of the number of classes in case of faces94 data

The Faces94 dataset contains images of 150 subjects. Each of these subjects is represented by 20 images. Hence, this image dataset gives us an opportunity to compare the performances

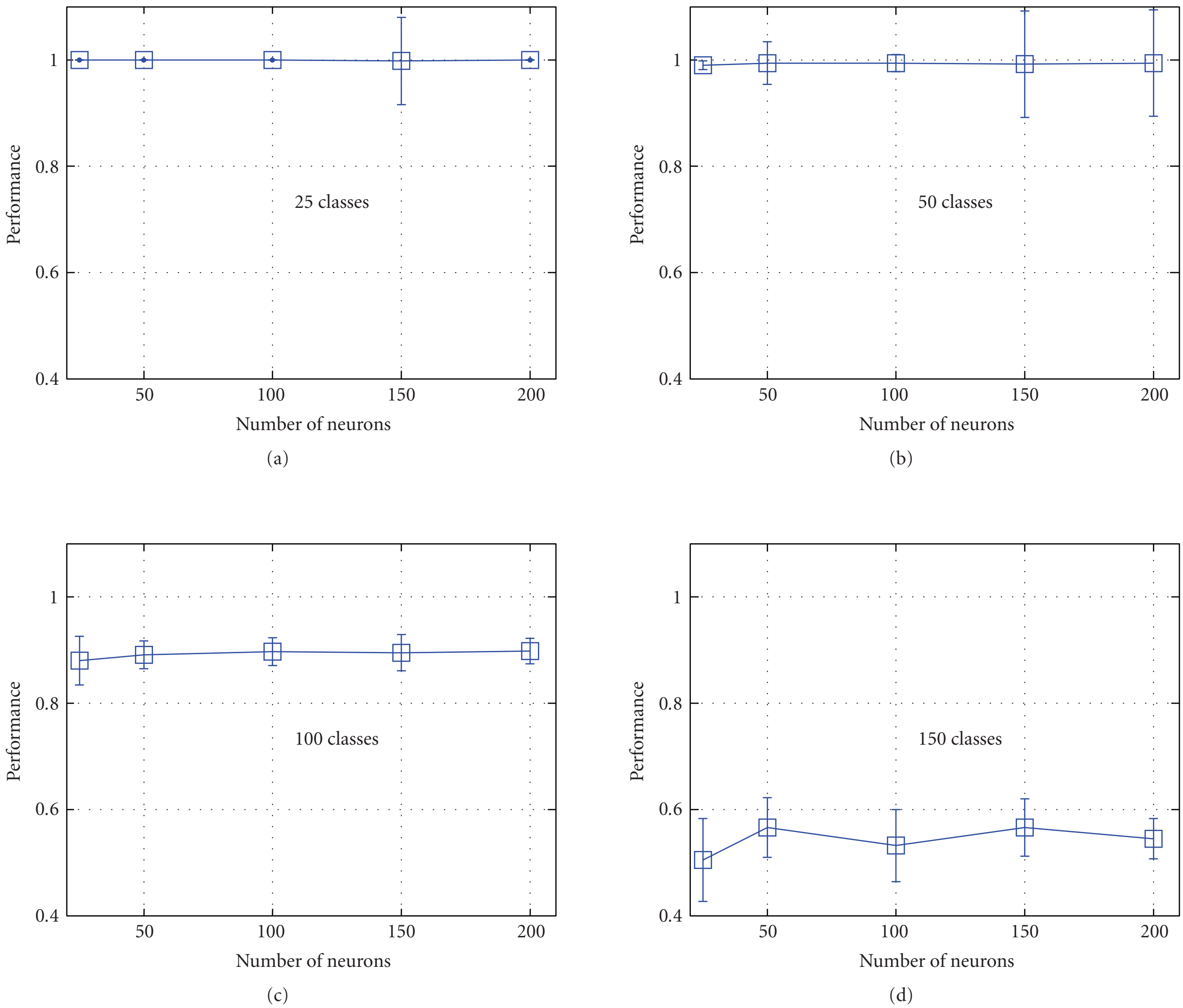


FIGURE 7: Performances of the multiclass recognition systems over the number of hidden neurons for 25, 50, 100, and 150 classes. Solid lines and bars are the mean and 2σ intervals, respectively.

of the proposed and multiclass recognition systems against different number of classes (subjects). In our experiments, we vary the number of classes between 25 and 150 as depicted in Figure 6.

From this figure, we can see that when the number of classes varies between 25 and 50, the performance of both systems in terms of predictive accuracy is close to maximal. However, when the number of classes increases, the performance of the multiclass system declines while the performance of the pairwise system remains near to maximal.

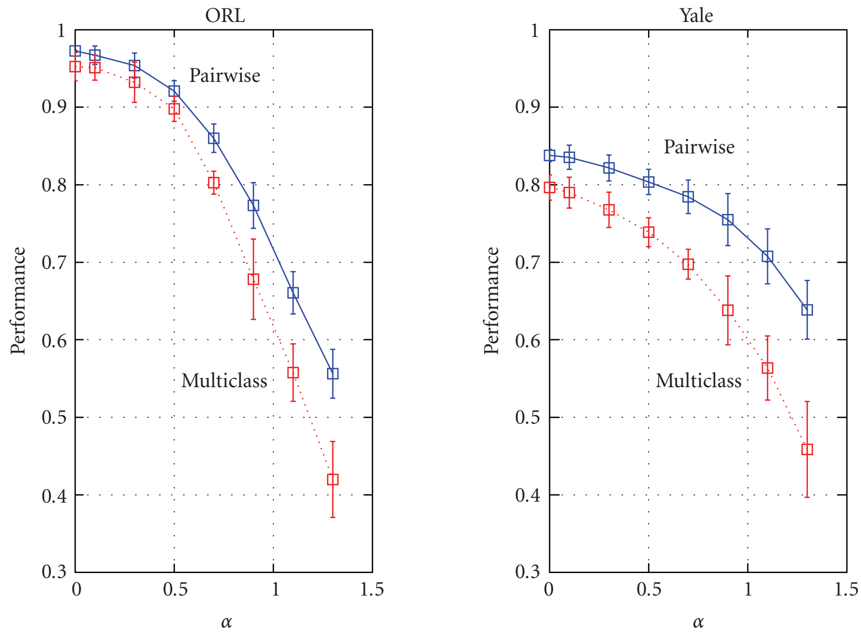
In these experiments, the best performance of the multiclass system was obtained with 100 hidden neurons. Figure 7 shows the performance of the multiclass system versus the numbers of hidden neurons under different numbers of classes.

From this figure, we can observe first that the number of hidden neurons does not contribute to the performance much. In most cases, the best performance is achieved with 100 hidden neurons.

4.6. Robustness to noise in ORL and Yale datasets

From our observations, we found that the noise existing in face image data can seriously corrupt class boundaries, making recognition tasks difficult. Hence, we can add noise of variable intensity to face data in order to examine the robustness of face-recognition systems. The best way to make data noisy is to add artificial noise to principal components representing face-image data. An alternative way is to add such noise directly to images. However, this method affects only the brightness of image pixels not the class boundaries locations.

For this reason in our experiments we add artificial noise to the principal components representing the ORL and Yale



(a) (b)

FIGURE 8: Performance of the pairwise and multiclass recognition systems over noise levels alpha. Solid lines and bars are the mean and 2σ intervals, respectively.

TABLE 1: Performance of the pairwise and multiclass recognition systems over the number of data samples. The performances are represented by the means and 2σ intervals.

Classification system	Number of data samples per class			
	50	100	150	200
Pairwise	0.965 ± 0.0103	0.981 ± 0.0124	0.983 ± 0.012	0.976 ± 0.0253
Multiclass	0.796 ± 0.157	0.812 ± 0.1485	0.807 ± 0.157	0.774 ± 0.1515

data in order to compare the robustness of the proposed pairwise and multiclass recognition systems. The performances of the pairwise and multiclass recognition systems over different noise levels are shown in Figure 8.

From this figure, we can see that for alpha ranging between 0.0 and 1.3, the proposed pairwise system outperforms the multiclass system. For instance, for alpha = 0.0, a gain in the performance is 2.0% on the ORL and 4.0% on the Yale datasets. For alpha = 1.1, the gain becomes 10.2% and 14.1%, respectively.

5. CONCLUSION

In order to reduce the negative effect of noise, corruptions, and variations in face images, we have proposed a pairwise neural-network system for face recognition. We assumed that the use of such classification scheme can improve the robustness of face recognition. Such assumption has been made on the base of our observations that the boundaries between pairs of classes are corrupted by noise much less than the boundaries between all the classes. High density of data can also make the recognition task difficult for multiclass systems.

We have compared the performances of the proposed pairwise and multiclass neural-network systems on the synthetic data as well as on the real face images. Having estimated the mean values and standard deviations of the performances under different levels of noise in the image data and different numbers of classes and samples per subject, we have found that the proposed pairwise system is superior to the multiclass neural-network system.

Thus, we conclude that the proposed pairwise system is capable of decreasing the negative effect of noise and variations in face images. Clearly, this is a very desirable property for face recognition systems when the robustness is of crucial importance.

ACKNOWLEDGMENTS

The authors are thankful to Dr. Libor Spacek from the University of Essex as well as to the AT&T Laboratories Cambridge for making the Faces94 and ORL face-image data available online. The authors are also thankful to the anonymous reviewers for their constructive comments.

REFERENCES

[1] S. Y. Kung, M. W. Mak, and S. H. Lin, *Biometric Authentication: A Machine Learning Approach*, Pearson Education, Beijing, China, 2005.

[2] C. Liu and H. Wechsler, "Robust coding schemes for indexing and retrieval from large face databases," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 132–137, 2000.

[3] A. S. Tolba, A. H. El-Baz, and A. A. El-Harby, "Face recognition: a literature review," *International Journal of Signal Processing*, vol. 2, no. 2, pp. 88–103, 2005.

[4] T. Hastie and R. Tibshirani, "Classification by pairwise coupling," in *Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems 10*, pp. 507–513, Denver, Colo, USA, 1998.

[5] V. Schetinin, J. Schult, B. Scheidt, and V. Kuriakin, "Learning multiclass neural-network models from electroencephalograms," in *Knowledge-Based Intelligent Information and Engineering Systems*, vol. 2773 of *Lecture Notes in Computer Science*, pp. 155–162, Springer, Oxford, UK, 2003.

[6] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision (ACV '94)*, pp. 138–142, Sarasota, Fla, USA, December 1994.

[7] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[8] "Faces94 face image data," http://cswww.essex.ac.uk/mv/allfaces/faces94.html.