

Face Recognition Committee Machines: Dynamic Vs. Static Structures

Ho-Man Tang, Michael R. Lyu, Irwin King
Department of Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong
{hmtang, lyu, king}@cse.cuhk.edu.hk

Abstract

We propose a Dynamic Face Recognition Committee Machine (DFRCM) consisting of five well-known state-of-the-art algorithms in this paper. In previous work, we have developed a static committee machine which outperforms all the individual algorithms in the experiments. However, the weight for each expert in the committee is fixed and cannot be changed once the system is trained. We propose a dynamic architecture on the committee machine which uses the input face image in the gating network to improve the overall performances. In addition, we adopt a feedback mechanism on the committee machine to adjust the weight of an individual algorithm according to the performance of the algorithm. Detail experimental results of different algorithms and the committee machine is given to demonstrate the effectiveness of the proposed system.

1 Introduction

In recent years, the committee machine, an ensemble of estimators, has proven to give more accurate results than the use of a single predictor. Jacobs et al. [11] proposed that the training data be partitioned into different subsets and various networks be trained on each subset. The basic idea is to train a committee of estimators and combine the individual predictions to achieve improved generalization performance. Different approaches are proposed by researchers within the last ten years [3] [2] [4] [18]. In [15] there exists two types of structure:

1. **Static Structure:** This is generally known as an ensemble method. Input data is not involved in combining the committee experts. Examples include ensemble averaging and boosting.

2. **Dynamic Structure:** Input is directly involved in the combining mechanism that employs an integrating unit to adjust the weight of each expert according to the input.

Recently, researchers have applied the committee machine in various fields. Su and Basu used the gating network on the image deblurring problem [16]. In face recognition, Gutta et al. used an ensemble of Radial Basis Function (RBF) network and a decision tree in the face processing problem [8] [9]. Huang et al. formulated an ensemble of neural networks for pose invariant face recognition [10].

In previous work [17], we presented a Face Recognition Committee Machine (FRCM) consisting of five experts including Eigenface [19], Fisherface [1], Elastic Graph Matching (EGM) [12], Support Vector Machine (SVM) [20] and Neural Networks. It was a static committee machine and it fused the knowledge acquired by the experts to arrive at a unified decision. The results in [17] demonstrated that the committee machine works in improving the accuracy.

In this paper, we propose a Dynamic FRCM which is an enhancement of the previous static one. We employ the gating network in a dynamic architecture which accepts input images and gives its corresponding weights for different experts. The rest of the paper is organized as follows. Section 2 describes the Static Face Recognition Committee Machine. Section 3 describes the Dynamic Face Recognition Committee Machine. Section 4 presents the feedback mechanism and describes the overall algorithm. Section 5 reports experimental results and Section 6 provides a conclusion.

2 Static FRCM

Our FRCM adopts the static structure with five well-known experts: 1) Eigenface, 2) Fisherface, 3)

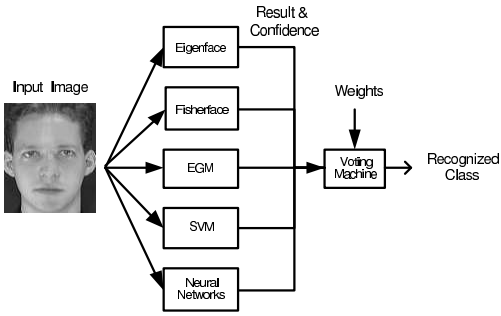


Figure 1. Static Structure

EGM, 4) SVM and 5) Neural Networks. All the algorithms are well-known algorithms in the field [21]. A brief review of the algorithms is given in [17]. The overview of the structure is shown in Fig. 1. Input image is sent to the five experts for recognition. Each expert gives its result r , together with a confidence c for the result, to the voting machine for assembling the results. We apply confidence values as weighted votes in the voting machine to avoid low confidence result of individual expert from affecting the final result. We adopt different approaches to find the results and the associated confidence to cope with different nature of the experts.

- **Eigenface, Fisherface and EGM:** We employ K nearest-neighbor classifiers, where five nearest training set images with the test image are chosen. The final result for expert i is defined as the class j with the highest votes v in J classes among the five results:

$$r_i = \arg \max_j (v(j)), \quad (1)$$

where its confidence is defined as the number of votes of the result class divided by K , i.e.,

$$c_i = \frac{v(r_i)}{K}. \quad (2)$$

- **SVM:** As SVM was originally developed for two-class classification, we use "one-against-one" approach in our implementation to recognize multiple classes. We construct $J C_2$ (i.e., $\frac{J(J-1)}{2}$) SVMs to recognize a test image in J different classes. The image is tested against each SVM and the class j with the highest votes in all SVMs is selected as

the recognition result $r(i)$. The confidence is defined like Equation 2 with $J - 1$ (the maximum number of votes for a class) instead of K .

- **Neural Networks:** We choose a binary vector of size J for the target representation. The target class is set to 1 and the others are set to 0. The class j with output value closest to 1 is chosen as the result and the output value is chosen as the confidence.

The weights w in static FRCM are evaluated in cross-validation testing for different algorithms under ORL [14] and Yale face database. We take the average performance for the algorithms in the experiments and normalize them to ensure the weights are positive and between 0 and 1 by an exponential mapping function:

$$w_i = \frac{\exp(p_i)}{\sum_{i=1}^5 \exp(p_i)}, \quad (3)$$

where p_i is the average performance of expert i . The use of weights in the voting machine further reduces the chance for an expert who performs poorly on average from affecting the ensemble result even if it has high confidence on a particular result. The voting machine assembles the results by calculating the score s of each class as follows:

$$s_j = \sum_{i=1}^5 w_i * c_i, \forall j \in r_i. \quad (4)$$

We define the score in such a way that only experts with high performance on average and high confidence on the result would take the most significant score in the final decision.

3 Dynamic FRCM

We propose the use of dynamic structure of committee machine on face recognition. The input is involved in the final decision of weights for different experts. In previous work, we find that the static structure has two major drawbacks:

- **Fixed weights under all situations** - The weight for the experts are fixed no matter what images are given to the committee machine. Different algorithms may perform variously under different situations. For example, global comparison methods like Eigenface and Fisherface achieve good results under normal situation but perform badly under strong lighting variations. It is undesirable to fix the weights for the experts under all situations.

- **No update mechanism for weights** - The weight for the experts cannot be updated once the system is trained. For a real-time verification system, we would like to train the system continuously. In the static structure, it does not contain a mechanism to update the weights.

To overcome these drawbacks, we develop a gating network which includes a neural network to accept input images and assign a specific weight for individual expert. We use several face databases [7] [13] including faces with various scale, rotation, lighting to model different situations. Instead of using a single set of average performance in static structure, we train the committee machine with the databases and store the experts' performances on each database separately. Table 1 shows the databases used and approximate number of images for each database.

Table 1. List of Face Databases

| Face Database | Number of Images |
|---------------|------------------|
| ORL | 400 |
| Yale | 165 |
| AR Face | 2500 |
| Umist | 560 |
| HRL | 1370 |
| Feret | 1200 |

In the training phase, we select 50 images from each database randomly as the training set data to train the neural network. The dimension of input images are first reduced by PCA to 50 for efficient training and recognition. Similar to the neural network training method mentioned before, we use binary vector of size 6 for target representation.

Fig. 2 shows the structure of Dynamic Face Recognition Committee Machine. In the recognition phase, test image is sent to five experts and the gating network. Each expert gives its result and confidence and the gating network determines the image's face database and return the corresponding weights of different experts to the voting machine. The voting machine takes the maximum score class as the finalized result.

4 Feedback Mechanism

For the dynamic structure, each expert is trained independently on different face databases in the training phase. Expert's performance, a measure for the weight for the expert, is then determined in the testing phase, which is defined as:

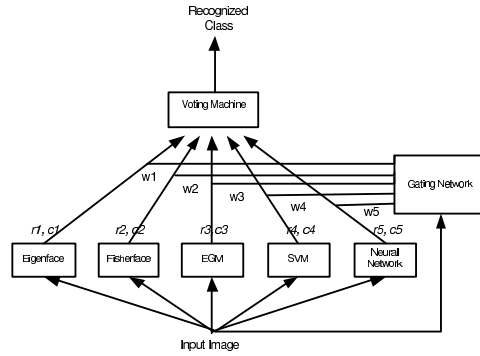


Figure 2. Dynamic Structure

$$p_{i,j} = n_{i,j}/t_{i,j}, \quad (5)$$

where $n_{i,j}$ is the total number of correct recognition and $t_{i,j}$ is the total number of trail for expert i on face database j .

1. Initialize $n_{i,j}$ and $t_{i,j}$ to 0 $\forall i, j$.
2. Train each expert i on different database j .
3. while TESTING
 - (a) Determine j for each test image.
 - (b) Recognize the image in each expert i
 - (c) if $t_{i,j} \neq 0$ then Calculate $p_{i,j}$.
 - (d) else Set $p_{i,j} = 0$
 - (e) Calculate $w_{i,j}$.
 - (f) Determine ensemble result.
 - (g) if FEEDBACK then Update $n_{i,j}$ and $t_{i,j}$.
4. end while

Figure 3. Overall mechanism

We propose a feedback mechanism to update the weights for the experts continuously in order to make the committee machine more secure and suitable for the verification purpose. In a face authentication system, a person with an identity is given to the system to verify whether the claimed identity matches the person or not. It is desirable to have a feedback mechanism to keep training the system in achieving low false acceptance and low false rejection. The overall mechanism

is proposed in Fig. 3.

The total number of correct recognition and trail are first initialized to 0 for all experts on each database. In the training phase, we train the experts on each database using the training set data. In the testing phase, the gating network determines the database j of the test image. Each expert then recognizes the image and gives its result and confidence. If the total number of trail for the experts is not 0, we can calculate the performance $p_{i,j}$ by Equation (5). Otherwise, we set the performance to 0. The performance are normalized by Equation (3) to obtain the weight $w_{i,j}$ for each expert. We can determine the ensemble result with the given weights, results and confidences from the gating network and experts in the voting machine. We update the total number of correct and trail for the current testing if feedback is required. In the recognition phase, the procedure is similar to the testing phase excepts that we do not have feedback as the correct result for the test image is unavailable.

5 Experimental Results

To evaluate the performance for both the SFRCM and DFRCM and the individual experts, we tested these systems with two well-known face database: 1) The ORL Database of Faces and 2) Yale Face Database. Experimental results including individual experts' performance and the proposed committee machines are given to demonstrate the effectiveness of the proposed approaches. In the testing, we employ leaving-one-out cross validation method to provide a thorough comparison. For a given sample of n images in a class, a classifier is trained using $(n - 1)$ images in that class and tested on the remaining images. The test repeats n times, each time training a classifier with leaving-one-out. A brief description of the databases and the partition of the data set are given below:

5.1 Database

The first experiment is evaluated on ORL face database from AT&T Laboratories Cambridge. The database contains 400 images, including 40 distinct people, each with 10 images that vary in position, rotation, scale and expression. Fig. 4 shows a snapshot of 4 individuals.

The second experiment is evaluated on Yale face database from Yale University. The database contains 165 images, including 15 distinct people, each with 11 images that vary in both expression and lighting. Fig. 5 shows a snapshot of 4 individuals.



Figure 4. Snapshot of ORL database



Figure 5. Snapshot of cropped Yale database

5.2 Dataset

For the static structure, we use all $(n - 1)$ images in the training set and test the remaining image to evaluate the performance. The weights used for the voting machine are based on the overall average performance of the experts in the testing.

For the dynamic structure, we partition the data into three sets: 1) training set, 2) testing set and 3) recognition set. Training set data (6 images) are used to train individual expert. Testing set data (3 images in ORL and 4 images in Yale) are used to measure the performance of each expert. The recognition set data (the remaining image) are used to evaluate the performance of all the experts and the committee machine.

5.3 Results

In the ORL Face testing, Table 2 and 3 list the results of DFRCM and SFRCM respectively. For DFRCM and SFRCM, they reach better performance (97.0% and 98.8%) than all the other experts in the testing. Table 4 shows the detail of the underlying data (including individual expert result and confidence) in Image Set 1 and 7 to show how the committee machine works. In Set 7, none of the experts has 100% accuracy but FRCM achieves it. The result also demonstrates

Table 2. DFRCM ORL Result

| S | Eigen | Fisher | EGM | SVM | NN | DFRCM |
|-------|-------|---------------|---------|---------------|--------------|---------------|
| 1 | 82.5% | 90.0% | 90.0% | 92.5% | 97.5% | 92.5% |
| 2 | 85.0% | 100.0% | 72.5% | 100.0% | 97.5% | 100.0% |
| 3 | 87.5% | 100.0% | 57.5.0% | 100.0% | 92.5% | 100.0% |
| 4 | 75.0% | 92.5% | 67.5% | 95.0% | 87.5% | 100.0% |
| 5 | 72.5% | 97.5% | 72.5% | 90.0% | 87.5% | 95.0% |
| 6 | 82.5% | 90.0% | 70.0% | 97.5% | 87.5% | 95.0% |
| 7 | 80.0% | 92.5% | 57.5% | 92.5% | 90.0% | 97.5% |
| 8 | 77.5% | 87.5% | 67.5% | 95.0% | 87.5% | 95.0% |
| 9 | 75.0% | 90.0% | 62.5% | 97.5% | 90.0% | 100.0% |
| 10 | 85.0% | 97.5% | 72.5% | 95.0% | 92.5% | 95.0% |
| P_i | 80.3% | 93.8% | 69.0% | 95.5% | 91.0% | 97.0% |

Table 3. SFRCM ORL Result

| S | Eigen | Fisher | EGM | SVM | NN | SFRCM |
|-------|-------|---------------|-------|---------------|--------------|---------------|
| 1 | 92.5% | 100.0% | 90.0% | 95.0% | 92.5% | 95.0% |
| 2 | 85.0% | 100.0% | 72.5% | 100.0% | 95.0% | 100.0% |
| 3 | 87.5% | 100.0% | 85.0% | 100.0% | 95.0% | 100.0% |
| 4 | 90.0% | 97.5% | 70.0% | 100.0% | 92.5% | 100.0% |
| 5 | 85.0% | 100.0% | 82.5% | 100.0% | 95.0% | 100.0% |
| 6 | 87.5% | 97.5% | 70.0% | 97.5% | 92.5% | 97.5% |
| 7 | 82.5% | 95.0% | 75.0% | 95.0% | 95.0% | 100.0% |
| 8 | 92.5% | 95.0% | 80.0% | 97.5% | 90.0% | 97.5% |
| 9 | 90.0% | 100.0% | 72.5% | 97.5% | 90.0% | 100.0% |
| 10 | 85.0% | 97.5% | 80.0% | 95.0% | 92.5% | 97.5% |
| P_i | 87.5% | 98.3% | 77.8% | 97.8% | 93.0% | 98.8% |

that with the use of confidence and weight function, poor results from some experts would not affect the ensemble result significantly.

In the Yale Face testing, Table 5 and 6 list the results of DFRCM and SFRCM respectively. Similar to the ORL testing, both committee machines has better performance (81.8% and 86.1%) than all the algorithms. Note that in Table 6, the performance for all individual drops significantly in Set 4 and 7. The main reason for the non-satisfactory results is due to the fact that Yale database contains variations in strong left and right lighting in Set 4 and 7 respectively. For algorithms taking the whole image as input like Eigenface, the accuracy would drop significantly because the lighting would greatly affect the pixel values.

5.4 Discussion

From the results, we note that Fisherface and SVM achieves better accuracy than other algorithms as both of them has better classification ability in general cases. These show that Fisherface and SVM are the best among the five algorithms.

In ORL test, both SFRCM (97.0%) and DFRCM (98.8%) outperform all the individual algorithms. The reason for the improvement is that different experts

Table 4. Detail SFRCM Result

| | | Recognized Class/Confidence | | | | | |
|---|----|-----------------------------|--------|--------|--------|--------|--------|
| S | I | Eigen | Fisher | EGM | SVM | NN | SFRCM |
| 1 | 0 | 15/0.4 | 0/0.6 | 20/0.2 | 15/1.0 | 23/0.4 | 15/0.3 |
| | 34 | 14/0.6 | 34/0.8 | 28/0.4 | 14/1.0 | 14/0.6 | 14/0.5 |
| 7 | 25 | 27/0.4 | 27/1.0 | 10/0.2 | 25/1.0 | 25/0.5 | 25/0.3 |
| | 34 | 26/0.4 | 18/0.6 | 34/0.4 | 34/1.0 | 34/0.4 | 34/0.4 |

Table 5. DFRCM Yale Result

| S | Eigen | Fisher | EGM | SVM | NN | DFRCM |
|-------|--------------|---------------|--------------|---------------|---------------|---------------|
| 1 | 40.0% | 73.3% | 46.7% | 93.3% | 60.0% | 100.0% |
| 2 | 73.3% | 93.3% | 66.7% | 86.7% | 86.7% | 86.7% |
| 3 | 73.3% | 86.7% | 86.7% | 86.7% | 93.3% | 86.7% |
| 4 | 26.7% | 40.0% | 13.3% | 26.7% | 40.0% | 40.0% |
| 5 | 93.3% | 100.0% | 86.7% | 100.0% | 93.3% | 100.0% |
| 6 | 86.7% | 93.3% | 86.7% | 86.7% | 93.3% | 86.7% |
| 7 | 26.7% | 40.0% | 66.7% | 20.0% | 26.7% | 40.0% |
| 8 | 66.7% | 93.3% | 80.0% | 93.3% | 86.7% | 93.3% |
| 9 | 80.0% | 93.3% | 60.0% | 100.0% | 93.3% | 93.3% |
| 10 | 73.3% | 53.3% | 46.7% | 66.7% | 46.7% | 73.3% |
| 11 | 93.3% | 86.7% | 46.7% | 100.0% | 100.0% | 100.0% |
| P_i | 66.7% | 77.6% | 62.4% | 78.2% | 74.5% | 81.8% |

contain various wrong data sets in the testing. The wrong data sets are disjoint among the experts since they are different in nature. When combining the results of all the experts, we may obtain a correct ensemble result. In Yale test, we notice that lighting has critical effects on recognition. All the algorithms perform badly in leftlight and rightlight image. To further improve the committee machine, we should include an expert for lighting variation like illumination Cone [6] [5] to tackle the light illumination problems.

Note that SFRCM is slightly poor than DFRCM, this may causes by the fact that DFRCM contains more training set data than that of SFRCM (9 vs. 6 images). As the Eigenface, Fisherface and EGM are implemented using 5 nearest-neighbor classifier, Only 6 images may not be sufficient to determine a correct recognition class. However, as the results shown, the committee machines still improve the overall performance than the individual expert.

6 Conclusion

In this paper, we propose a novel Dynamic Face Recognition Committee Machine. Based on previous work on the Static Face Recognition Committee Machine, we introduce the use of a gating network which accepts input image to identify the situation that the image is taken. Different weights are given to each expert according to the identification result from the gating network. We use different face databases taken

Table 6. SFRCM Yale Result

| S | Eigen | Fisher | EGM | SVM | NN | SFRCM |
|-------|---------------|---------------|--------------|---------------|---------------|---------------|
| 1 | 53.3% | 93.3% | 66.7% | 86.7% | 73.3% | 93.3% |
| 2 | 80.0% | 100.0% | 53.3% | 86.7% | 86.7% | 100.0% |
| 3 | 93.3% | 100.0% | 80.0% | 100.0% | 93.3% | 100.0% |
| 4 | 26.7% | 26.7% | 33.3% | 26.7% | 26.7% | 33.3% |
| 5 | 100.0% | 100.0% | 80.0% | 100.0% | 100.0% | 100.0% |
| 6 | 86.7% | 100.0% | 86.7% | 100.0% | 93.3% | 100.0% |
| 7 | 26.7% | 40.0% | 40.0% | 13.3% | 26.7% | 33.3% |
| 8 | 86.7% | 93.3% | 93.3% | 100.0% | 93.3% | 100.0% |
| 9 | 86.7% | 100.0% | 73.3% | 100.0% | 100.0% | 100.0% |
| 10 | 86.7% | 66.7% | 33.3% | 73.3% | 66.7% | 86.67% |
| 11 | 100.0% | 100.0% | 66.7% | 93.3% | 93.3% | 100.0% |
| P_i | 75.2% | 83.6% | 64.2% | 80.0% | 77.6% | 86.1% |

from various situations to model the scenario. We also propose a feedback mechanism on the committee machine to keep updating the weights of the experts.

The success of both the DFRCM and SFRCM has been demonstrated on the results of ORL and Yale database. In our experiments, DFRCM achieves 97.0% and 81.8% accuracy while SFRCM achieves 98.8% and 86.1% accuracy, on ORL and Yale face database, respectively. These improvements in accuracy are greater than other state-of-the-art face recognition algorithms.

7 Acknowledgments

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. The Chinese University of Hong Kong 4360/02E & 4351/02E).

References

- [1] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, volume 19, pages 711–720, July 1997.
- [2] L. Breiman. Bagging predictors. Technical report, Department of Statistics, University of California, Berkeley, California 94720, September 1994.
- [3] L. Breiman. Some infinite theory for predictor ensemble. Technical Report 577, Statistics Department, UC Berkeley, August 2000.
- [4] Y. Freund and R. E. Schapire. A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*, 15(5):771–780, September 1999.
- [5] A. S. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose. In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, volume 23, pages 643–660, June 2001.

- [6] P. Georghiades, A.S. and Belhumeur and D. Kriegman. From few to many: generative models for recognition under variable pose and illumination. In *Proc. Fourth IEEE International Conf. on Automatic Face and Gesture Recognition, 2000*, pages 277–284, 2000.
- [7] D. B. Graham and M. Allinson, Nigel. Characterizing virtual eigensignatures for general purpose face recognition. *Face Recognition: From Theory to Application*, pages 446–456, 1998.
- [8] S. Gutta, J. Huang, B. Takacs, and H. Wechsler. Face recognition using ensembles of networks. In *IEEE Proc. 13th International Conference on Pattern Recognition*, volume 3, pages 50–54, August 1996.
- [9] S. Gutta, J. R. J. Huang, P. Jonathon, and H. Wechsler. Mixture of experts for classification of gender, ethnic origin, and pose of human faces. In *IEEE Trans. on Neural Networks*, volume 11, pages 948–960, July 2000.
- [10] F. J. Huang, H.-J. Zhang, T. Chen, and Z. Zhou. Pose invariant face recognition. In *IEEE Proc. Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 245–250, 2000.
- [11] R. A. Jacobs, M. I. Jordan, J. N. Steven, and E. H. Geoffrey. Adaptive mixtures of local experts. In *Neural Computation*, volume 3, pages 79–87, 1991.
- [12] M. Lades, J. Vorbruggen, C. von der Malsburg, R. P. Wurtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. In *IEEE Trans. on Computers*, volume 42, pages 300–311, March 1993.
- [13] A. M. Martinez and R. Benavente. The ar face database. Technical Report 24, Computer Vision Center, June 1998.
- [14] F. Samaria and A. Harter. Parameterisation of a stochastic model for human face identification. In *Proc. IEEE Workshop on Application of Computer Vision, Sarasota FL*, December 1994.
- [15] M. Su and M. Basu. Gating improves neural network performance. In *IEEE Proc. International Joint Conference on Neural Networks*, volume 3, pages 2159–2164, 2001.
- [16] M. Su and M. Basu. Gating improves neural network performance. In *Proc. IEEE Conf. on IJCNN '01*, volume 3, pages 2159–2164, 2001.
- [17] H. M. Tang, M. R. Lyu, and I. King. Face recognition committee machine. In *Proc. IEEE International Conf. on Acoustics, Speech, and Signal Processing*, volume 2, pages 837–840, 2003.
- [18] V. Tresp. Committee machines. In *Handbook on neural Network Signal Processing*. CRC Press, 2001.
- [19] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 586–591, 1991.
- [20] V. Vapnik. *Statistical Learning Theory*. John Wiley and Sons, Inc., 1998.
- [21] J. Zhang, Y. Yong, and M. Lades. Face recognition: eigenface, elastic matching, and neural nets. In *Proc. IEEE*, volume 85, pages 1423–1435, 1997.