An RCE-based Associative Memory with Application to Human Face Recognition

Xiaoyan Mu, Mehmet Artiklar, and Mohamad H. Hassoun
Department of Electrical and Computer Engineering
Wayne State University
Detroit, MI 48202
hassoun@brain.eng.wayne.edu

Paul Watta
Department of Electrical and Computer engineering
University of Michigan-Dearborn
Detroit, MI 48128
watta@umich.edu

Abstract

In this paper we construct an associative memory model based on the restricted Coulomb energy (RCE) network. We propose a simple architecture and training algorithm for this RCE-based associative memory. We study the capacity of this memory model on the practical problem of human face recognition. In this case, capacity is described by two measures: the ability of the system to correctly identify known individuals, and the ability of the system to reject individuals who are not in the database. Experimental results are given which show how the performance of the system varies as the size of the database increases—up to 1000 individuals.

1. Introduction

In the associative memory problem, we are given a set of fundamental associations to be stored in memory:

$$\mathbf{DB} = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_m, \mathbf{y}_m)\}$$

The task is to design a system so that the following 3 conditions are satisfied:

1. When presented with input $\mathbf{x}_i$, the system produces output $\mathbf{y}_i$.

2. When presented with a noisy version of $\mathbf{x}_i$, the system still produces $\mathbf{y}_i$ at the output.

3. When presented with an input $\mathbf{x}$ that is very dissimilar to all of the known inputs $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ (that is, $\mathbf{x}$ is a garbage input), the system should reject the input.

Various models of associative memory have been proposed in the neural network literature [1]. Simple models like the single layer Hopfield network are (to a limited extent) able to satisfy conditions 1 and 2, but not 3. Hence these networks have limited practical use as associative memory because they cannot distinguish between meaningful patterns and pure noise.

The restricted Coulomb energy (RCE) network is typically used for pattern recognition problems. In this paper, we explore its use as an associative memory. An attractive feature of the RCE network is that it has a natural rejection mechanism which can be used as the basis for rejecting patterns with very low signal-to-noise ratio. In the remainder of this paper, we describe a simple architecture and learning algorithm for the RCE-based associative memory and then show how it performs for the task of human face recognition. We measure the ability of the system to correctly identify known individuals, as well as its ability to reject unknown individuals.

2. The Model

In this section we describe an RCE-type neural network which operates as an associative memory and is suitable for use in a face recognition system. We will assume that the given database **DB** contains K different examples of each association:

$$\mathbf{DB} = \{(\mathbf{x}_{mk}, y_m) : m = 1, 2, \ldots, M;\ k = 1, 2, \ldots, K\}$$

In the case of face recognition, $\mathbf{x}_{mk}$ is the kth sample of individual m and y_m is the name of individual m. For example, we may include a collection of image samples of each individual, showing different facial expressions or lighting conditions, etc. [2].

2.1 Architecture

Typically, the RCE network is a unit allocating system whereby the number of units is increased as training proceeds [3, 4]. We will simplify the training by just allocating one hidden unit for each prototype image in the database. The network structure is shown in Figure 1.

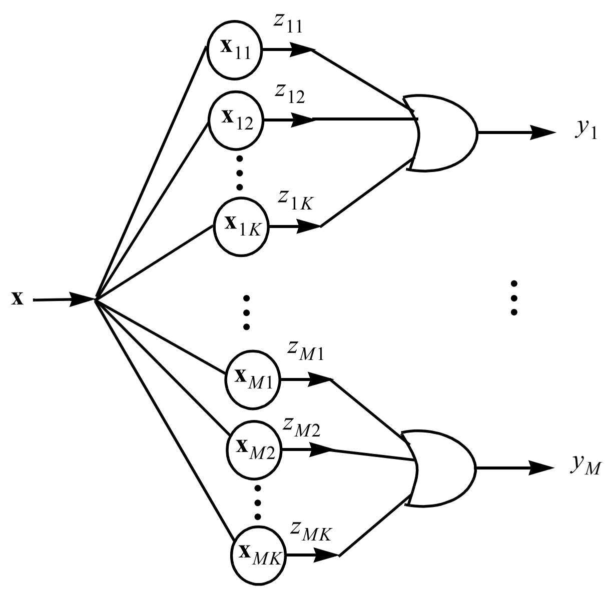

Figure 1. Architecture of the RCE network.

Associated with each hidden layer neuron in Figure 1 is a threshold or radius which is used to determine whether the input is sufficiently close to the prototype so that the unit should fire. The output z_{mk} of the *mk*th hidden unit (with radius r_{mk}) is given by:

$$z_{mk} = \begin{cases} 1 & \text{if } \; d(\mathbf{x}_{mk}, \mathbf{x}) \le r_{mk} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Clearly, since the outputs y_m are obtained by ORing together the outputs of the hidden units $z_{m1}, z_{m2}, \ldots, z_{mK}$, then $y_m = 1$ if any of these hidden units fire; otherwise $y_m = 0$. For an input key of say the *m*th person, we want the output vector **y** to consist of all 0's except for a single 1 at the *m*th position.

The system will reject the input in the following cases: (1) More than 1 output unit fires, and (2) none of the output units fire.

2.2 Training

For the *mk*th hidden unit, the radius parameter r_{mk} is computed as follows. First, we compute the (between-class) distances between $\mathbf{x}_{mk}$ and all the other images in the database (excluding images of person *m*):

$$d_{ij} = \{d(\mathbf{x}_{mk}, \mathbf{x}_{ij}) : i = 1, \ldots, M; i \ne m; j = 1, \ldots, K\} \tag{2}$$

And then choose the minimum such distance

$$d_{i^*j^*} = \min\{d_{ij}\} \tag{3}$$

The number $d_{i^*j^*}$ represents the largest radius that unit *mk* can have, because at that radius, another (incorrect) unit (namely, the i^*j^* unit) will also fire. Hence we set the radius r_{mk} as a fraction of that radius:

$$r_{mk} = \alpha d_{i^*j^*} \tag{4}$$

where $0 < \alpha < 1$. The selection of α is application dependent. For example, a relatively small value of α is needed for high security systems where it is critical that the number of misclassifications and number of false positive matches are minimal. Of course setting α too low results in a large number of rejections. For the face database used here, we found that a value of $\alpha = 0.8$ yields a good compromise.

2.3 Shift Invariance

Noise is always present in any practical image capturing system [2]. There are several ways to make the RCE-based associative memory less prone to small shifts (translations) and rotations in the image. For example, in the elastic band matching, the distance between the input and each of the database images is computed on a rectangular grid of points and neighbors of the grid points [5]. The collection of grid points which yields the smallest distance is chosen (deformed template).

We will employ the simple and computationally efficient method outlined in [6] to provide invariance to small amounts of image shift. The method involves optimizing the relative translation of one image with respect to another so that the resulting distance is minimized (or similarity maximized). The method is an iterative process and uses a greedy algorithm of locally moving in the best direction at each step [6].

3. Methods

A detailed discussion of the construction of the CNNL face database that is used in these experiments can be found in [7]. Briefly, the database consists of 1000 different men and women of various ages (between 15 and 65 years old). The images were collected in a university setting and hence the database contains samples from a variety of ethnic backgrounds. In order to minimize tilt, rotation, shifting, scaling, and changes in illumination, the images were collected in a laboratory setting. A head brace was used to position the subject's head in front of the camera, but note that no attempt was made to normalize the position of the eyes, etc. Hence, these images do not require sophisticated preprocessing. In fact, the only preprocessing done on the images is intensity normalization.

For each person, two groups of 4 training images were snapped, giving a total of 8 training images per person. The first group of 4 images all show a blank facial expression. The second group of 4 images show different facial expressions: blank, smile, angry, and surprised. Two test images were also collected for each person: a test image where the subject shows a blank facial expression, and a test image where the subject gives an arbitrary or unusual expression to try to fool the recognition system.

Both the database and test images were snapped at a dimension of 82×115 and stored as 8-bit gray scale (256 levels). The 82×115 images were cropped to a size of 72×72 (using a fixed cropping position) in order to eliminate the hair and forehead of the person from the image. Figure 2 samples of the 72×72 images for two of the subjects in the database.

In this paper, we use a rigorous testing methodology to study how the performance of the RCE associative memory varies as the size of the database is increased. We will assess system performance by performing 2 different types of experiments:

1. **Correct classification experiments**. Measure the ability of the system to correctly identify known individuals. There are 3 important measures to consider here: percent correct, percent rejected, and percent misclassified. Of course since these quantities sum to 100%, only two need to be reported. We will report percent rejected and percent misclassified.

2. **False positive experiments**. Measure the ability of the system to reject individuals who are not part of the database. Here the measure of merit is the number of false positive matches (i.e., incorrectly identified as a known individual).

4. Results

For these experiments, **DB** will consist of the 4 facial expression images (blank, smile, angry, and surprised) of each of the M known individuals (so $K = 4$). Since there's a total of $4M$ images in the database, there's $4M$ hidden units in the RCE network.

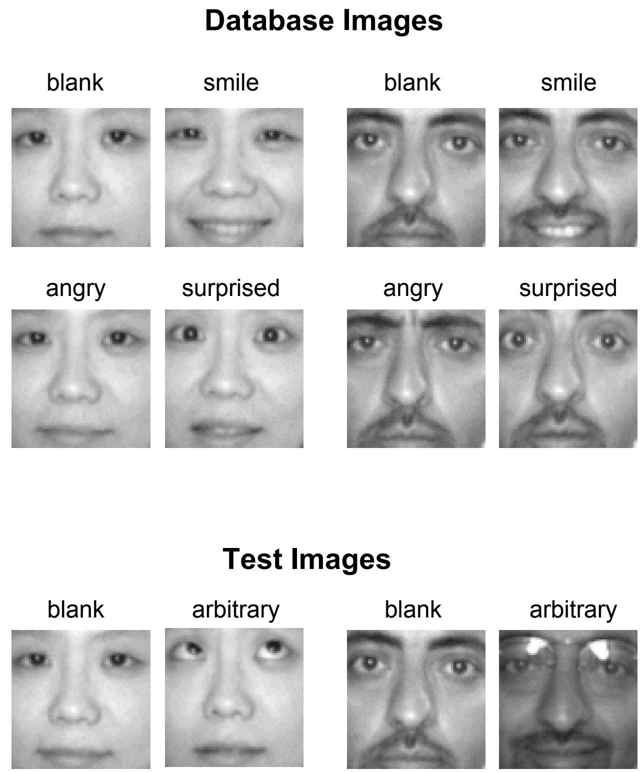

Figure 2. Samples of the 72×72 database and test images for two different people (the 4 additional blank images are not shown). In total, there are 1000 individuals in the database.

4.1 Correct Classification Experiments

In the training phase, the database images are used to determine the radius r_{mk}, $m = 1, 2, \ldots, M$; $k = 1, 2, 3, 4$, for each of the hidden units. After training, the system was tested for both correct classification performance and false positive performance. To see how the performance scales with increasing number of subjects in the database, we looked at the performance of the system as the number of people in the database varied from 100 to 1000 individuals.

Table 1 shows the results of the correct classification experiments for the RCE-based associative memory for both the blank and arbitrary expression test images. Note that each number in Table 1 (except for the $M = 1000$ case) is an average over 5 different trials. In each trial, a different database was (randomly) chosen from the available 1000 individuals.

The results show that the algorithm scales well and for the blank expression test images gives an error rate of 0.1% (or less) with a reject rate of less than 5%. As expected, for the arbitrary expression test images, the system rejects a much larger number. However, there are no classification errors on the arbitrary expression test set.

Num People M	Blank		Arbitrary	
	Reject	Error	Reject	Error
100	3.4	0	28.4	0.4
200	2.7	0	31.8	0
300	2.7	0	33.4	0
400	3.4	0	36.5	0
500	3.5	0	36.2	0
600	4.0	0.1	37.9	0
700	3.5	0.1	39.3	0
800	4.2	0.1	39.3	0
900	4.2	0.1	40.7	0
1000	4.5	0.1	41.4	0

Table 1. Correct classification results for the blank and arbitrary expression test sets as the number of people in the database *M* increases.

4.2 False Positive Experiments

Next we ran false positive experiments and tested the system with 200 randomly chosen individuals who are not present in the database to see how well the system rejects these images. In this case, we used all available training and test images (10 images total) for each test person. Note that we could not perform this experiment on the $M = 1000$ database due to lack of additional images, and could only test with 100 people for the $M = 900$ database.

The results of the false positive experiments are shown in Table 2. Curiously, the number of false positive matches decreases as the number of people in the database increase, and is only 0.4% for $M = 900$. The reason for this is that as more images are added to the database, the hidden layer thresholds necessarily become tighter as Equations (2)-(4) are computed over more images. Another reason is that as more and more images are added to the database, it is more likely that more than 1 output unit will fire. But, by our simple decision rule, if more than 1 output fires, we reject the input.

4.3 The Effect of Radius Size α

For all the results given above, a threshold scaling parameter of $\alpha = 0.8$ was used. Figure 3a shows how the rejection and false positive results vary as a function of α.

Num People M	FP Matches Error
100	2.9
200	1.6
300	1.1
400	1.0
500	1.2
600	0.7
700	0.6
800	0.4
900	0.4

Table 2. Percentage of false positive matches as the size of the database increases.

Of course to get good false positive performance, we require a very small value for α. But setting α too low causes a lot of rejections (in the correct classification experiments) because the condition for a hidden unit to fire (Equation 1) is too stringent. On the other hand, setting α too high also causes a lot of rejections because it is more likely that more than 1 output unit will fire.

To analyze the rejection performance of the system, we measured the number of rejections due to the fact that *no hidden units were activated*, as well as the number of rejections due to the fact that *more than* 1 *output unit was activated*. Figure 3b shows the result of this experiment on a database of $M = 500$ individuals. The sum of both of these quantities is the number of images rejected by the system.

5. Discussion

A plot of both the classification (rejection) and false positive results is shown in Figure 4. From this figure, we see that the general trend in rejection performance increases roughly linearly (after $M = 200$). We did not plot the number of misclassified images, since it was so low: there was at most 1 misclassified image for each test. However, it is expected that as *M* increases, the number of misclassified images will also begin to rise. A larger database is needed in order to perform those experiments.

When collecting the images for the CNNL database, even though we tried to limit the amount of variation in scale, rotation, intensity, we did not completely eliminate such effects. An analysis shows that some of the test images that were rejected by the RCE network had some or all of these types of distortion. For example, in one case, the subject is looking straight for all the database images; however, in the test image, the subject is looking slightly to the left. In other cases, there are obvious intensity variations between database and test images.

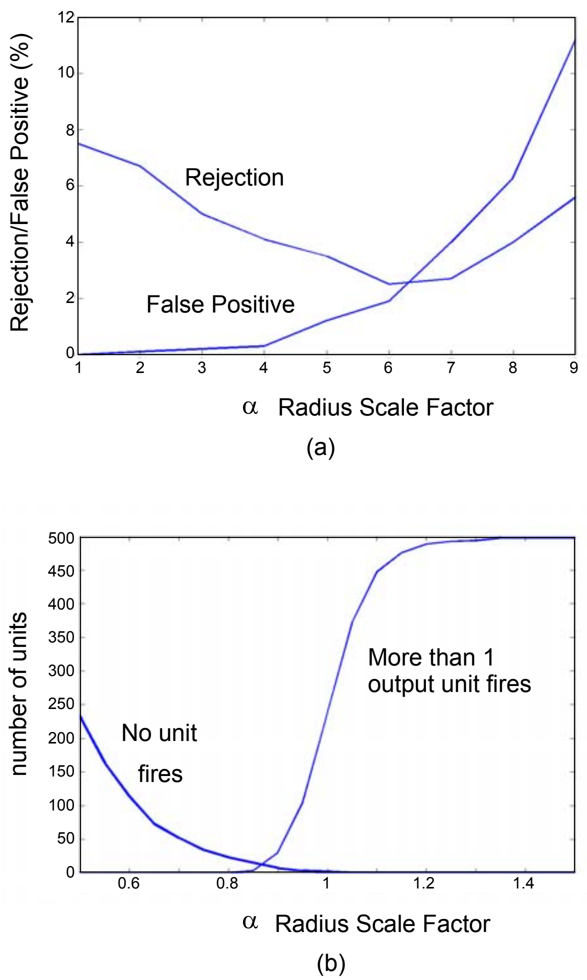

Figure 3. (a) Rejection and false positive performance as a function of α . (b) Number of times no unit fires and more than 1 unit fires as a function of α . The sum of these is the number of images which are rejected by the RCE network.

For the proposed system, training simply involves computing inter-class distances among the database images. Hence, the training scales linearly with the number of people in the database.

An important design consideration for face recognition systems is how easy it is to add and delete associations from the memory set. For the given RCE network, it is easy to delete individuals from the database—simply delete the hidden units allocated to that person (or else shrink the radius to 0). To add individuals, one simply allocates additional hidden units for each new image sample.

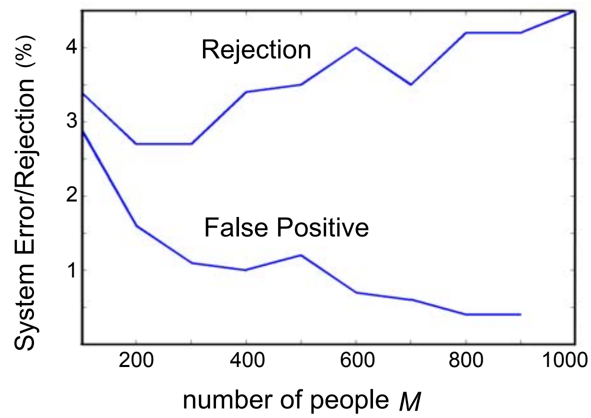

Figure 4. Results of the rejection performance and false positive performance of the system as a function of database size.

In terms of adding individuals to the database, the fact that the number of false positive matches decreases as M increases (see Figure 4), is a very nice property and allows us to use a modular approach for constructing and training large databases.

For example, suppose we start with a $M = 400$ person database. Now suppose we want to add an additional 400 people to the database. Rather than retraining on the whole 800 person system, we can just formulate and train a separate 400-person network to handle the new patterns. Then the outputs of the original network and the new network can be combined with a simple decision: If both networks reject the input image or both networks classify the input image, then the image is rejected. However, if one of the networks classifies the pattern and the other network rejects it, then we accept that classification. Figure 5 shows a schematic diagram of this modular design.

Note that we expect to use such a modular approach for large size databases ($M = 1000$ and above) where the number of false positive matches are sufficiently low. We did, however, test the idea on an 800-person database. In this case, two 400-people RCE networks were independently trained and then the outputs combined with the simple decision rule described above. The results are shown in Table 3. The correct classification results show remarkable similar performance. The false positive performance is a bit worse for the two 400 person network, but not by much.

Such a modular design approach has obvious practical advantages for large-scale systems where the stored associations must be updated in bulk periodically. For example, consider a face recognition application in a university setting, where every year a new group of freshman students must be added to the system.

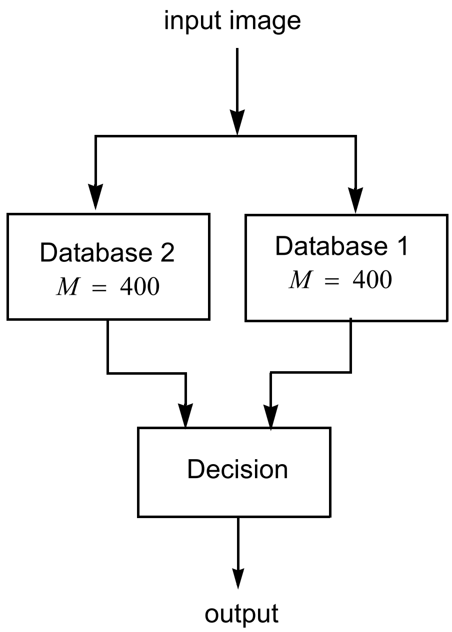

Figure 5. Block diagram of the modular construction of a face recognition system. Here a $M = 800$ person database is constructed as two independent 400-person databases along with a decision network.

Test Set	One 800 Person Database		Two 400-person Databases	
	Reject	Error	Reject	Error
Blank	4.2	0.1	4.6	0.4
Arbitrary	39.3	0	37.6	0.1
False Positive	-	0.4	-	0.65

Table 3. Performance of the modular RCE network (two 400-person databases) compared to the full network (a single 800 person database).

5. Summary and Conclusion

The RCE-based associative memory developed here has a natural rejection mechanism and hence is able to identify meaningful patterns and reject all patterns with low signal-to-noise ratio. We showed that, assuming the face images are suitably preprocessed, the RCE-based associative memory can achieve good performance on the human face recognition problem. The performance of the system on face images acquired under less constrained conditions is currently under investigation.

References

1. Hassoun, M. H., ed. (1993). *Associative Neural Memories: Theory and Implementation*. Oxford University Press, New York.

2. Chellappa, R., Wilson, C., and Sirohey, S. (1995). “Human and Machine Recognition of Faces: A Survey,” *Proceedings of the IEEE*, **83**(5), 705-740.

3. Reilly, D. L., Cooper, L. N., and Erlbaum, C. (1982). “A Neural Model for Category Learning,” *Biological Cybernetics*, **45**, 35-41.

4. Hasegawa, A., Shibata, K., Itoh, K., Ichioka, Y., and Inamura, K. (1996). “An Adaptive Neural Network: Application to Character Recognition on X-ray Films,” *Neural Networks*, **9**(1), 121-172.

5. Duc, B. and Fischer, S. (1999). “Face Authentication with Gabor Information on Deformable Graphs,” *IEEE Transactions on Image Processing*, **8**(4), 504-516.

6. Artiklar, M., Hassoun, M., and Watta, P. (1999). “Application of a Postprocessing Algorithm for Improved Human Face Recognition,” *Proceedings of the IEEE International Conference on Neural Networks*, IJCNN-1999, July 10-16, 1999, Washington, DC., Paper #JCNN 2166.

7. Watta, P., Artiklar, M., Masadeh, A., and M. H. Hassoun, (2000). “Construction and Analysis of a Database of Face Images which Requires Minimal Preprocessing,” *Proceedings of the IASTED Conference on Modeling and Simulation,* MS-2000, May 15-17, 2000, Pittsburg Pennsylvania, 465-469.