# CBA generated receptive fields implemented in a Facial expression recognition task

Jose M. Jerez[1], Leonardo Franco[2], and Ignacio Molina[3]

[1] scuela Técnica Superior de Ingeniería en Informática
Departamento de Lenguajes y Ciencias de la Computación
Universidad de Málaga
jja@lcc.uma.es
[2] Center for Computational Neuroscience
Department of Experimental Psychology
University of Oxford
leonardo.franco@psy.ox.ac.uk
[3] Escuela Técnica Superior de Ingeniería en Telecomunicación
Departamento de Tecnología Electrónica
Universidad de Málaga
aimc@dte.uma.es

E

**Abstract.** Biologically inspired receptive fields are used to process input facial expressions in a modular network architecture. Local receptive fields constructed with a modified Hebbian rule (CBA) are used to reduce the dimensionality of input images while preserve some topological structure. In a second stage, specialized modules trained with backpropagation classify the data into the different expression categories. Thus, the neural net architecture includes 4 layers of neurons, that we train and test with images from the Yale Faces Database. A generalization rate of 82.9% on unseen faces is obtained and the results are compared to values obtained with a PCA learning rule at the initial stage.

## 1 Introduction

Face perception is a very important component of human cognition. Faces are rich in information about individual identity, but also about mood and mental state, being accessible windows into the mechanisms governing our emotions. Facial expression interactions are relevant in social life, teacher-student interaction, credibility in different contexts, medicine, etc. Face expression recognition is also useful for designing new interactive devices offering the possibility of new ways for humans to interact with computer systems.

From a neurophysiological point of view face recognition appears to be very important. Experiments both in monkeys and humans show the existence of dedicated areas in the brain where neurons respond selectively to faces ([1–3]). Also it has been shown that complex visual processing related to discrimination of faces is a very rapid task that can be completed in approximately 100 msec suggesting the involvement of a mainly feed-forward neural mechanism [4].

J. Mira and J.R. Álvarez (Eds.): IWANN 2003, LNCS 2686, pp. 734-741, 2003.
© Springer-Verlag Berlin Heidelberg 2003

In this work we construct a modular neural network including two parts, trained in different ways: first, a hidden layer of neurons having the task of developing receptive fields, with the aim of reducing the dimensionality of the data, to be further classified by the second stage of specialized modules, sometimes called "experts", trained with backpropagation.

Neural-based models have been proposed in the last decades to imitate the perceptual capabilities of simple cells in striate cortex of biological systems [5–8], and some have been tested in natural scenarios [9, 10]. Stimulating a single neuron model with natural images, PCA [11] learning rule was shown to develop receptive fields (RFs) similar to those found in visual cortex in the early experiments of Hubel and Wiesel [12, 13]. The resulting RFs contained excitatory and inhibitory regions arranged in a preferred orientation. The emergence of these regions are related to long term potentiation (LTP) and depression (LTD) of the learning rule. According to the Hebbian postulate, PCA rule incorporates heterosynaptic LTD but not homosynaptic LTD, both described in the nervous system by different authors [14, 15]. The CBA rule proposed in this work to perform the first processing stage incorporates both types of synaptic competition and it is derived from the one proposed in [16] for neural assemblies formation, with the difference that incorporates a decay term.

The use of modular neural network architecture rather than fully connected ones seems to be a simple and effective solution to complicated tasks and also have the advantage of better generalization properties, reduced training times, and adaptability [17, 18]. Modular networks have been used successfully in several tasks such as speaker verification, face identification, time series prediction, etc. [19–21], and are also very useful tools for exploring hypothesis about brain function [22, 23].

Different systems have been constructed to deal with facial expressions ([24] for an interesting review), but few of them use a neural network approach (see for instance [25] and references therein). For example, in [26] a feedforward network with PCA input encoding of some facial features (eyes and mouth) is trained to classify emotions; Lisetti et al. [25] constructed a backpropagation network to classify the degree of expressiveness of faces; and Franco et al. [27] use an unsupervised algorithm to reduce the dimensionality of the input data and a self-organizing process to form the receptive fields. This work continues to explore the potential of neural networks to perform this kind of task, trying to respect some biological constraints implemented by the CBA rule, and using the capabilities of modular systems to obtain better classification rates.

## 2 The Database of Images

The Yale Face Database [28] contains 165 gray scale images in GIF format of faces from 15 male individuals of different races and containing different features (moustache, glasses, etc.) We use a subset of the database that consists of 14 subjects displaying 4 facial expressions: neutral, happy, sad and surprised faces. The images were cropped to obtain input images 8 pixels width by 24 pixels height covering a portion of the face located on the left side. (See Figure 1). The images were centered taking the tip of the nose as reference and some light illumination correction was applied to a couple of images; both operations were carried out by a human observer. The resulting images were transformed to pgm 8-bit gray scale format, ready to be fed into the network after a linearly scaled transformation of pixel intensity into the interval $[0, 1]$.

Figure 1 shows a sample of the different expressions displayed by one of the subjects. In the leftmost image the area of the face cropped and used as input is shown.

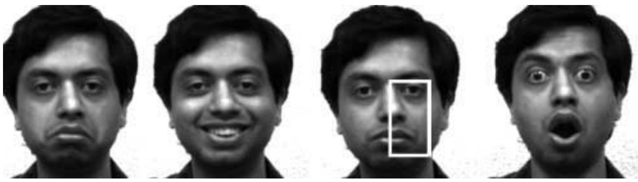

**Fig. 1.** Sample subject showing the four full face expressions (neutral, happy, sad and surprised). The white rectangle inside the rightmost figure corresponds to the area cropped and used as input for the neural network.

## 3 Network architecture

The network architecture is similar to that proposed in the work of Franco et al. [27], and it consists of a 4 layer modular neural structure composed of sigmoidal units. The input layer has 192 units corresponding to the 24x8 pixels of the area cropped from the original images. Every input neuron transmits information through a single hebbian weight, projecting to a specific neuron in the first hidden layer, selected according to a self-organized process. Thus, one has at this level a new reduced representation of the images, expressed by the activity of 48 neurons, that preserves some topological aspects of the original input. The whole network architecture is shown in figure 2, where at the top we show a sample input image followed by the structure of the receptive fields corresponding to the first hidden layer neurons.

After this unsupervised compression the network splits into three modules corresponding to the expressions different from the neutral face: happy, sad and surprised. The structure of the modules could depend on the emotion they specialize in; in the case we consider here they have all the same type of architecture: one hidden layer fully connected with one output unit. There is a difference in the number of hidden neurons belonging to each modules since the recognition of happy and surprised faces is much easier than the recognition of sad ones, a fact that was previously known from experiments both with humans and computers [22]. It was necessary to put 4 hidden neurons for sad faces while 3 neurons were enough for happy and surprised ones. In this way the output of the whole network has 3 neurons that should be all OFF when a neutral face is presented, while when a face displaying an emotion is shown, the corresponding module output unit should be ON.

### 3.1 Receptive fields formation process

As we mentioned before, the first layer of weights is organized according to the dynamics of both CBA and PCA learning rules, with the aim of obtaining receptive fields of $2 \times 2$ pixels from the input images.

The neuron model consists of a vector $\mathbf{x}$ of inputs, a vector $\mathbf{w}$ of synaptic weights, and a scalar output $y$, given by $y = \mathbf{w} \cdot \mathbf{x}$, that represents the neuron activity.

In a previous work [16] we proposed a new correlational learning rule (BA, for bounded activity) that formed stable neural attractors in a recurrent network. The CBA learning rule is essentially a modification of the BA rule in which an extra term to implement the heterosynaptic LTD has been incorporated. Thus, the resulting synaptic modification equation for the CBA rule is a Hebbian-type learning rule with an specific form of stabilization, defined as

$$\Delta w_i = \alpha \, x_i \, y \, (y - \tau)(\lambda - y) - \beta \, y \, w_i, \tag{1}$$

where $\alpha$ is the learning rate, $y$ is the neuron activity and $x_i$ is the input activity from the $i$-synapsis. The term $\lambda$ avoids the unbounded weight growing in a plausible way, and can be interpreted as a neuronal parameter representing the maximum level of activity at which the neuron might work. The CBA modification equation also defines a threshold $\tau$ that determines whether depression or potentiation occurs depending on pre- and postsynaptic activities. Finally, the parameter $\beta$ controls the heterosynaptic competition effect, such that the strength of synapses can change even in the absence of presynaptic activity to those synapses. All these parameters adopt positive values. The effect of this adaptation mechanism in the receptive fields formation process is that the graded response elicited after stimulus presentation leads the neural activity either to high or resting levels.

On the other hand, the PCA rule, proposed by Oja [11], is known to perform a principal component analysis of the input vectors, converging to the largest eigenvector, while normalization is ensured. The change in the weight values can be written as

$$\Delta w_i = \alpha \, y \, (x_i - y \, w_i) \tag{2}$$

## 4 Simulations results

As the amount of data available for training and testing is limited (14 subjects, 56 images), a cross-validation procedure was used [18]. In this procedure 13 out of the 14 available subjects are chosen to train the network and the 4 unseen

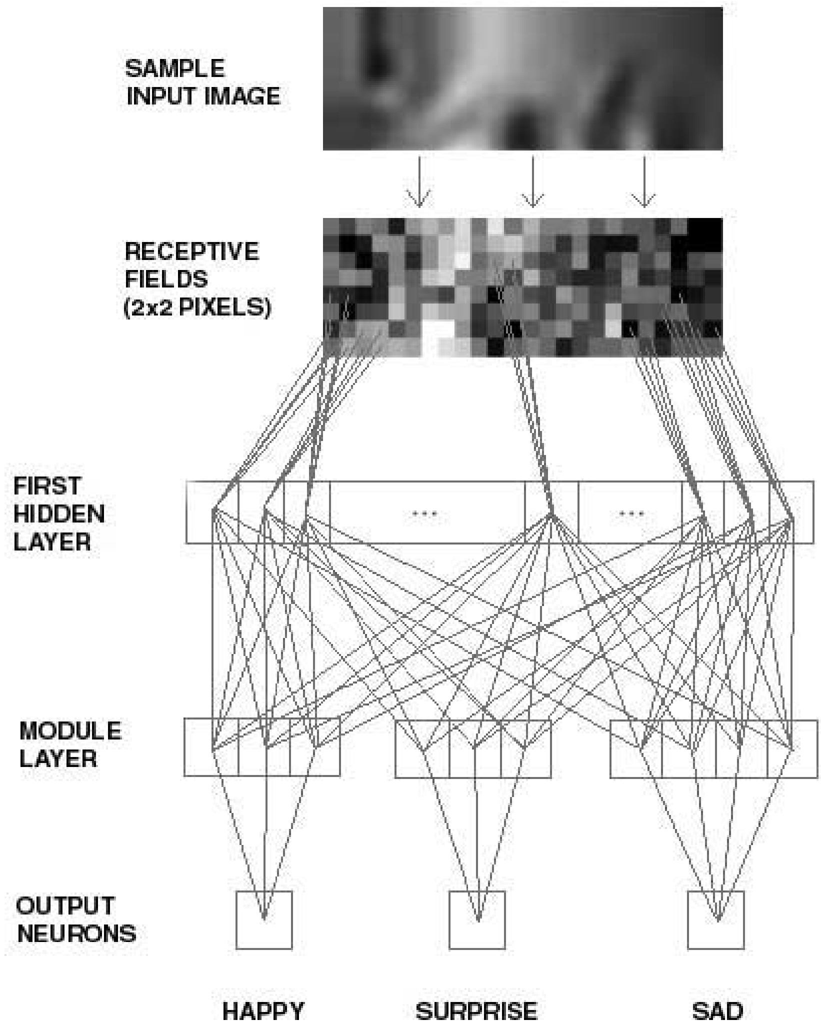


**Fig. 2.** Schematic structure of the network architecture used to perform facial expression recognition. The network has 192 inputs corresponding to the part of the face considered and being projected via hebbian weights to 48 neurons in the first hidden layer, with self-organized receptive fields. The modules, specialized in the different expressions: happy, sad and surprise, have a one hidden layer structure with an output that should be activated when a face displaying its corresponding expression is presented at the input. At the top, one of the input sample images is shown.

faces of the remaining subject are used to test the generalization ability of the system. The process is repeated 14 times, one for each subject being kept out of the training set, and averages values are taken.

The first layer of 192 weights, one for each input pixel, is trained with the CBA and PCA unsupervised learning rules. The rest of the weights, those belonging to the modules, were trained with the standard backpropagation algorithm (Hertz, Krogh & Palmer, 1993; Haykin, 1995). To prevent overtraining and to obtain a better generalization ability we monitor the training error on each input image, stopping the training on such image when the error was lower than 0.10. Since the backpropagation training is an on-line procedure, at the end of the training phase the average error per example was decreased to 0.02, approximately. All layers of weights were trained at the same time upon the presentation of an input image. Table 1 shows the parameters setting for the proposed neural architecture.

**Table 1.** Setting of parameters for the neural architecture (CBA, PCA and BP algorithms).

| | |
|---|---|
| Learning constant for BP algorithm, $\eta$ | 0.01 |
| Learning constant for unsupervised learning, $\alpha$ | 0.001 |
| Maximum level of activity, $\lambda$ | 1.0 |
| Threshold level, $\tau$ | 0.1 |
| Heterosynaptic competition term, $\beta$ | 0.000015 |
| Input values range, $x$ | [0.0, 1.0] |
| Weights initialization range, | $[0.1 \pm 0.001]$ |

The generalization error rates produced by the three modules specialized in different expressions for CBA and PCA learning rules are shown in table 2. We can observe that the generalization rate is similar for both rules, although the error is distributed uniformly among the three modules for CBA, whereas the recognition of the sad faces (more difficult to recognize than the others) is worst for PCA rule.

## 5 Conclusions and future work

We explore the generalization ability of a modular neural system to classify facial expressions. Using a mixed learning scheme composed by unsupervised-supervised training, we obtain a generalization ability on novel faces of 82.9%. The unsupervised processing used the CBA rule that incorporates an homosynaptic term that prevents an excessive growing of the weights and lead to a similar level of recognition rate than the PCA rule. An interesting feature of the results is that a similar performance is obtained for the different modules

**Table 2.** Generalization error rates for the modules, specialized in happy, sad and surprise faces, and for the whole net using first layer PCA and CBA rules in the formation of receptive fields. The generalization error is measured after the training procedure succeeds, when the training error per example turns out to be around 0.02.

| Expression Module | Error (PCA) | Error (CBA) |
|---|---|---|
| Happy | 0.057 | 0.057 |
| Sad | 0.086 | 0.057 |
| Surprise | 0.027 | 0.057 |
| Total | 0.170 | 0.171 |

specialized on different facial expressions, in contrast to the case when the PCA rule is implemented, and it may need further tests to be fully understood.

The advantage of using a modular approach is that it will permit the addition of new modules to recognize different expressions that could be trained separately.

Possible extensions of this work includes testing the system with a more extensive database, using the whole face as input, and studying the effect of changing the sizes of the receptive fields. It would also be desirable that the network itself should be capable of performing the identification of a face in a complex input image, permitting the use of the system in a real environment and potentially to be mounted on a robot.

## References


1. Kanwisher, N., McDermott, J., Chun, M.: The fusiform face area: A module in human extrastriate cortex specialized for face perception. Journal of Neuroscience **17** (1997) 4302–4311
2. Perret, D., Rolls, E., Caan, W.: Visual neurons responsive to faces in the monkey temporal cortex. Experimental Brain Research **47** (1982) 329–342
3. Desimone, R.: Face selective cells in the temporal cortex of monkey. Journal of Neuroscience **3** (1991) 1–8
4. Lehky, S.: Fine discrimination of faces can be performed rapidly. Journal of Cognitive Neuroscience **12** (2000) 848–855
5. Von der Malsburg, C.: Self-organization of orientation sensivity cells in striate cortex. Kybernetik **14** (1973) 85–100
6. Bienenstock, E., Cooper, L., Munro, P.: Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. Journal of Neuroscience **2** (1982) 32–48
7. Linsker, R.: From basic network principles to neural architecture: Emergence of orientation-selective cells. Proceedings of the National Academy of Science **83** (1986) 8390–8394
8. Miller, K.: A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between on-and off-center inputs. Journal of Neuroscience **14** (1994) 409–441

9. Law, C., Cooper, L.: Formation of receptive fields according to the bcm theory of synaptic plasiticity in realistic visual environment. Proceedings of the National Academy of Science **91** (1994) 7797–7801
10. Shouval, H., Liu, Y.: Principal component neurons in a realistic visual environment. Network **7** (1996) 501–515
11. Oja, E.: A simplified neuron model as a principal component analyzer. Mathematical Biology **15** (1982) 267–273
12. Hubbel, D., Wiesel, T.: Receptive fields, binocular interaction and functional architecture in the cats visual cortex. Journal of Physiology **160** (1962) 106–154
13. Hubbel, D., Wiesel, T.: Receptive fields and functional architecture of monkey striate cortex. Journal of Physiology **195** (1968) 215–243
14. Artola, A., Brcher, S., Singer, W.: Different voltage-dependent threshold for inducing long-term depression and long-term potentiation in slices of rat visual cortex. Nature **347** (1990) 69–72
15. Abraham, W., Goddard, G.: Asymmetric relationships between homosynaptic long-term potentiation and heterosynaptic long-term depression. Nature **305** (1983) 717–719
16. Vico, F., Jerez, J.: Stable neural attractors formation: Learning rules and network dynamics. Neural Processing Letters **to appear** (2003)
17. Franco, L., Cannas, S.: Generalization properties of modular networks implementing the parity function. IEEE Transactions on Neural Networks **12** (2001) 1306–1313
18. Haykin, S.: Neural Networks: A comprehensive foundation. IEEE Press (1994)
19. Bennani, Y.: Multi-expert and hybrid connectionist approach for pattern recognition: speaker identification task. International Journal of Neural Systems **5** (1994) 207–216
20. Dailey, M., Cottrell, G.: Organization of face and object recognition in modular neural networks models. Neural Networks **12** (1999) 53–1074
21. Petridis, V., Kehagias, A.: Predictive modular neural networks: Applications to time series. Kluwer Academic Publishers, Boston (1998)
22. Dailey, M., G.W., C., Adolphs, R.: A six-unit network is all you need to discover happiness. In: Proceedings of the 22th Annual Conference of the Cognitive Science Society. (2000)
23. Jacobs, R.: Nature, nurture, and the development of functional specializations: a computational approach. Psychonomic Bulletin and Review **4** (1997) 299–309
24. Fasel, B., Luettin, J.: Automatic facial expression analysis: a survey. Pattern Recognition **36** (2003) 259–275
25. Lisetti, C., Rumelhart, D.: Facial expression recognition using a neural network. In: Proceedings of the 11th International Florida Artificial Intelligence Research Society Conference, Menlo Park, CA (1998)
26. Padgett, C., Cottrell, G.: A simple neural network models categorical perception of facial expressions. In Mahwah, ed.: Proceedings of the 20th Annual Cognitive Science Conference, Madison, WI, Lawrence Erlbaum (1998)
27. Franco, L., Treves, A.: A neural network facial expression recognition system using unsupervised local processing. In Loncaric, S., Babic, H., eds.: ISPA-01 Proceedings of the 2nd IEEE R8-EURASIP International Symposium on Image and Signal Processing and Analysis. (2001) 628–632
28. Belhumeur, P., Kriegman, D.: The yale faces database. http://cvc.yale.edu/projects/yalefaces/yalefaces.html (1997)