ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 129-134

Modeling face recognition learning in early infant development

Francesca Acerra*, Yves Burnod* and Scania de Schonen**

*INSERM U483, Université Paris VI, 9 quai St Bernard, 75005 Paris, France
**CNRS (LDC), Université Paris V, 28 rue Serpente, 75006 Paris, France

Abstract. Face recognition development has been studied in experimental psychology, in the first month of life. These studies show that already at the age of 4 months the right hemisphere processes configural information, while the left hemisphere processes what is classically called local information. We have developped a neural model to understand how face recognition learning develops in early infancy. We propose a bayesian network based on local cellular properties of visual areas and on lateral-feedforward interactions in the cortex. The model reproduces the experimental data of the right hemisphere infant behavior, when tested with faces. We suggest that the bayesian neural networks and the biological properties of cortical areas may be a more general and useful instrument to understand human development.

Introduction

Face recognition is a sophisticated ability of humans. The development of this ability has been studied in the first months of life when the vision is limited. The experimental results have shown that :
1) the infants can see only low spatial frequencies [1],
2) some kind of face processing starts at birth.
Pascalis et al [2] have demonstrated that 3-day-old neonates recognize their mother's head but not their mother's face. When presented with the mother's face and a stranger's face, 3-day-old infants look longer at the mother. But this purely visual recognition vanishes when the mother and the stranger wears a scarf that masks the hairline and the line between forehead and hair. This changes such that by the end of the second month and during the third month of life that infants recognize their mother's face from their internal configuration even with a scarf.
Experimental results [3] suggest that two different systems for processing faces may develop at that time. One involves the right hemisphere and processes global information (as in adults), and the other one involves the left hemisphere and processes what is classically called local information. More precisely in the first year of life of the human infant the right hemisphere processes what we have called configural information and, more importantly, it does not differentiate precisely local information when embedded in a more complex pattern.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 129-134

We have developed a neural model to understand how face recognition and its learning develop in early infancy, and more precisely the property of the right hemisphere to process faces.

Hypothesis of the model

The model is a neural network (see Appendix) based on 4 properties of the infant visual system:
1) the spatial frequencies related to an immature retina: the network is therefore based on low spatial visual frequencies (from 0.02 to 0.5 c/d, 0.19 being the 3-day-old preferred frequency),
2) the tuning properties of neurons in area V1: the first processing is based on Gabor filters, as usually modelled,
3) the activity-dependent learning properties of both feedforward and lateral connections in cortical areas: the activities and learning rule of the network are based on a bayesian model of the lateral-feedforward interaction, consistently with neural and psychophysical data [4],
4) no interhemispheric connections seem to be functional before the age of 2 years; the model assumes that two perceptual systems, the right and the left, work differently and independently; we can thus focus on the specific property of the right hemisphere.

A natural learning

The learning procedures are designed to be as similar as possible with natural conditions:
1) stimuli are faces of 3 women each in 6 different orientations: inputs are thus 18 natural faces (black/white photo images), all pre-processed by a low pass filtering, to simulate the limitation concerning the high spatial frequencies; this first processing layer of the network decompose faces in Gabor wavelets with low spatial frequencies only,
2) for the first steps, the neural network learns only one face with 6 orientations, corresponding to natural conditions points of view. This corresponds to natural conditions of learning since infants have a bond with their mothers,
3) after, it "familiarizes" with the other two women, but the first one (the mother) is presented always two times more than the others. The network "overlearns" its "mother".

The test phase

We consider that the initial state of the network model the infant capacity at birth and that the final state, when the weights are stabilized, correspond to the infant

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 129-134

capacity at the age of 4 months. We can thus compare its capacity with the infant behaviors as measured in psychological experiments.
To test the results of the simulation we have thus built 6 other faces, 2 for each woman, with 2 different transformations which have been used in psychological experiments in babies (from the age of 4 to 10 months):
- "eye size" transformation: the eyes size is changed only (scale 1:4) : this transformation is easy to distinguish by the right hemisphere of 4 months babies,
- "eye shape" transformation: the eyes shape is changed only : this transformation is not distinguished by the right hemisphere of these infants.
During test phase we present to the network all 24 faces (the 18 previous learned + the 6 transformed) and we analyze the network ability to differentiate between the different stimuli.
We test the network at different ages to compare its evolution.
To have a measure of these differences we define the computed distance d_{IJ} after learning between two faces I and J, for each spatial frequency, as:

$$\forall C(I,J)\, d_{IJ}(f) = \sum_{k(f)} \begin{cases} 1 \text{ if } \left| I'_{k(f)} - J'_{k(f)} \right| >= s \\ 0 \text{ if } \left| I'_{k(f)} - J'_{k(f)} \right| < s \end{cases}$$

where s=0.01, k(f) is the index corresponding to the 30 positions related to the spatial frequency f, C(I,J) is the I, J pair, I' and J' are the network outputs corresponding to the inputs I and J.

Results

The network learns the category " face " : the distance between the different orientations of the same person and the distance between the different persons decrease. This corresponds to cerebral regions which are activated by faces, whatever the face which is presented. However we expect that this network should behave differently with the 2 transformations of faces tested at 4 months of age.

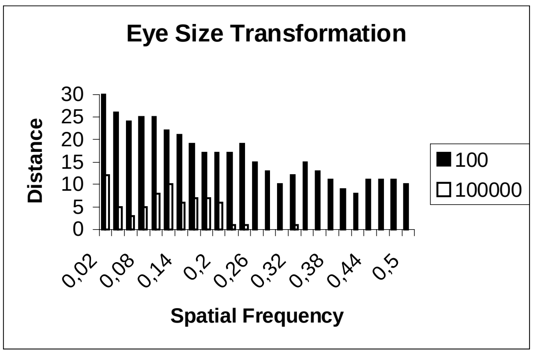

Fig. 1 – Network evolution for eye size distances

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 129-134

In Fig. 1 we show the network evolution of the 25 distances (each for each spatial frequency) between one face I and the same face with the "eyes size" transformation J : on the horizontal axis, are shown the 25 spatial frequencies of the input layer (see above), on the vertical axis, are shown the computed distances between I and J for each frequency. The value of the black bars shows the computed distances in the initial phase (age = 100 steps), while the white bars shows results when the network is stabilized (100000 steps, since there is no change from 10000 steps to 100000).
We can easy see that there is an important decrease of the distances for high frequencies, toward a nul value, while the distances for low frequencies decrease until a non nul value and then stabilize. This result showes that the network continues to make a difference between a face and its "eyes size" transformation: this is in perfect analogy with the infants right hemisphere behavior.

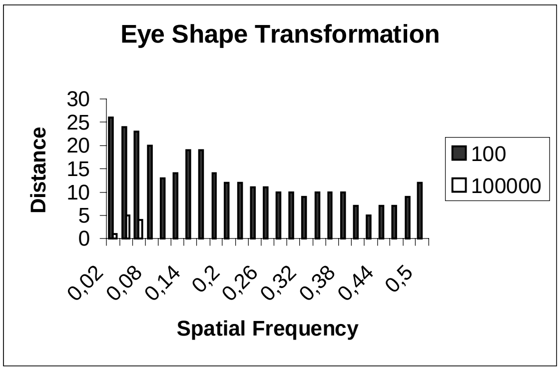

Fig. 2 – Network evolution for eye shape distances

Conversely, in Fig. 2, showing the evolution of the distance between one face I and the same face with the "eye shape" transformation K, we see an important decrease of the distances for all frequencies : the network does not discriminate between a face and its "eye shape" transformation, exactly like the right hemisphere of human babies.

Conclusion

The network learns to recognize the category “ face ”, and moreover distinguishes between a face and its "eye size" transformation: it makes, as it is called in psychology, a configural processing. Conversely, the network does not differentiate between a face and its "eye shape" transformation, and recognize these faces as the same: it is unable to process what is called, in psychology, local information.
These results are in full agreement with the experimental data on the right hemisphere infant behavior. Since the hypothesis are based on the cellular properties of the visual system, the results suggest that such properties, together with a bayesian learning process, are sufficient to explain the development of face recognition in

infants. Such network modeling may be an useful instrument to understand human development.

References

[1] Banks M. S., Salapatek P. (1981). Infant pattern vision: A new approach based on the contrast sensitivity function. Journal of Experimental Child Psychology, 31, 1-45.
[2] Pascalis O., Schonen S. de, Morton J., Deruelle C., Fabre-Grenet M. (1995). Mother's face recognition by neonates: a replication and an extension. Infant Behavior and Development, 18, 79-85.
[3] Schonen S. de, Mathivet E. (1989). First come, first served: a scenario about development of hemispheric specialization in face recognition during infancy. European Bulletin of Cognitive Psychology (CPC), 9, 3-44.
[4] Koechlin E., Anton J. L., Burnod Y. (1999). Bayesian inference in populations of cortical neurons: a model of motion integration and segmentation in area MT. Biological Cybernetics, 80 (1), 25-34.

Appendix

The network structure is shown in Fig. 3:

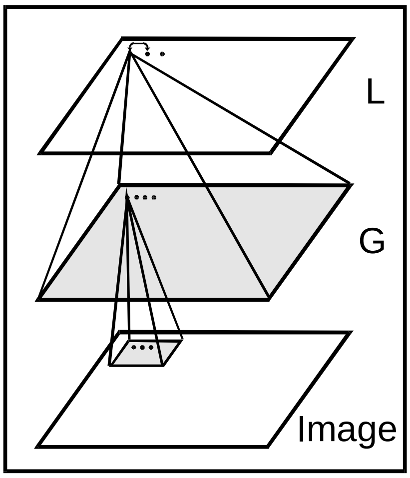

Fig.3 – The model structure.

The first layer (Image) represents the input face on the retina, a matrix of 112x90 black-white pixels. The second layer (G) represents the Gabor filtering s_i, an array of 25 spatial frequencies (from 0.02 to 0.5 c/d) x 30 positions. The third layer (L)

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 129-134

represents the lateral connected (by the weights r_{ij}) neurons x_j (25x30). We suggest that the probability of the population activity is maximized:

$$P\left\{\bigcup_j X_j\right\}^{(t+h,k)} = \sum_{ij} s_i(t) r_{ij} x_j(t) k$$

We obtain the following learning rules with the normalisation $\sum_i r_{ij} = 1$,

$$\frac{\tau_l}{r_{ij}} \frac{dr_{ij}}{dt} = \alpha\, x_j \left(s_i - \sum_i s_i r_{ij}\right)$$

where

$$\frac{dx_i}{dt} = \sum_j \alpha\, r_{ij} s_i(t) x_j(t) - x_i(t) \sum_j x_j(t)$$

x_i is the output of the network, an array of 25x30 neurons, each corresponding to one spatial frequency and one position.