HyberBF Networks for Gender Classification

R. Brunelli[1], **T. Poggio**[2,1]
[1] Istituto per la Ricerca Scientifica e Tecnologica
I-38050 Povo, Trento, ITALY
[2]Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139, USA

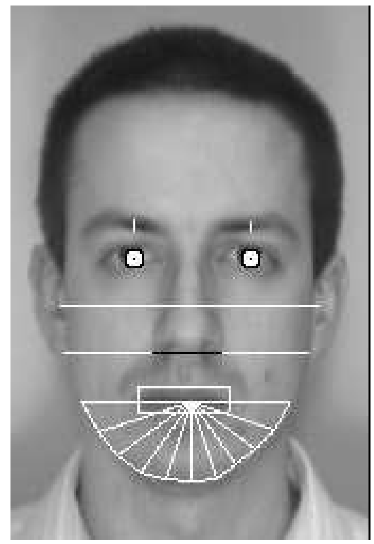
Abstract— **A set of geometrical features is extracted automatically from digitized pictures of frontal views of people without facial hair. This compact description is then used to train two competing HyperBF networks to classify according to gender. The results using a database of twenty males and twenty females show an average performance of 79% correct gender classification on images of new faces. Correct classification on vectors corresponding to new face images present in the training set but not used in the training phase rises to 86%. Preliminary experiments to assess human performance on the same set of grey level images give an average result of 90% which, while higher than network performance, suggests that peoples' performance is comparable. Interestingly, the HyperBF technique finds the relative weights of the different features and converges to prototypes of the male and female face that seem to exaggerate their difference, somewhat like caricatures do.**

1. Introduction

Faces allow people to establish, among other things, the gender of a person, his (her) age and, to a certain extent, emotions. In the current paper we address gender classification and will show how limited geometrical information accounts for correct sex attribution.

There are two main strategies for face recognition (and for object recognition in general): feature comparison and template matching. The former relies on a set of selected features which must be computed from an available image while the latter directly compares the appearance of a given instance with a reference image by means of a suitable metric. The first strategy, when feasible, works with a compact representation of the objects to be matched which are usually represented by low (as compared to the number of pixels of a template) dimensional vectors. The set of features used for recognition or classification is critical as it must capture the discriminating ones and give to each of them the correct weight.

In some recent work [4] the problem of face recognition and gender classification has been tackled using the internal representation of a compression network as unsupervised feature extractor and a (smaller) classification network taking as input the extracted features. Recent theoretical results [2] show that the internal representation of such a network is closely related to a Karhunen-Loewe expansion (see also [8], [13]) so that the work of Cottrell et al. should probably be considered as classified in the template matching category. In our paper we want to show how limited geometrical information (see Fig. 1 for the set of features) can give reasonable performance and possibly provide some insight into human mechanisms.



1	pupil to eyebrow separation
2	pupil to nose vertical distance
3	pupil to mouth vertical distance
4	pupil to chin vertical distance
5	eyebrow thickness
6	nose width
7	mouth width
8	bizygomatic breadth
9	bigonial breadth
10-15	six chin radii
16	mouth height

Fig. 1. Geometrical features (white) used in the face recognition experiments

2. Gender Classification

The inspection of a face allows us to establish, usually without much effort, the gender of the person we are looking at. It seems natural to mimic this ability with a computer program. The experiment we did is based on the use of a geometrical feature vector. In fact, the same vector extracted for recognition purposes in a previous paper [3] was used. The only difference is that the face description has been symmetrized (left and right eyebrow and chin information has been averaged) thereby reducing the dimensionality of the vector.

All of the features have been extracted automatically, from images whose rotation and scale was previously normalized (by automatically locating eyes). The paradigm we used is that of learning from examples, where a system learns to discriminate between males and females given a sufficient number of examples. The system we used is based on a classifier called Hyper Basis Function Network (see [10]).

Learning from examples can be regarded, whenever the inputs and output are expressible as numerical vectors, as the reconstruction of an unknown function from sparse data. From this point of view learning is equivalent to functional approximation. Hyper Basis Function Networks are a tool for multivariate function approximation and rest on a solid background of results in this field.

Before describing the networks used for gender classification let us briefly recall the fundamentals of the Hyper Basis Functions Network.

Radial Basis Functions can be regarded as a special case of Regularization Networks introduced in [10] as a general approximation technique that can be used in problems of learning from examples.

A scalar function can be approximated, given its value on a sparse set of points $\{\bar{x}_i\}$, by an expansion in radial functions:

$$F(\bar{x}) = \sum_{i=1}^{N} c_i G(\| \bar{x} - \bar{x}_i \|) \tag{1}$$

where $\| \cdot \|$ represents the usual Euclidean norm. The computation of the coefficients c_i rests on the invertibility of matrix $\mathbf{H}_{ij} = G(\| \bar{x}_i - \bar{x}_j \|)$ which has been proved (see Micchelli [9]) for functions such as:

$$G(r) = e^{-(\frac{r}{c})^2} \tag{2}$$

$$G(r) = (c^2 + r^2)^{\alpha}, \alpha < 1 \tag{3}$$

It is possible to use fewer radial functions than examples, i.e. data points. The resulting overconstrained system can be solved using a least square approach under the conditions of Michelli's theorem and proves to be useful when many examples are available [10].

Poggio and Girosi [11] have shown that the RBF technique is a special case of the regularization approach to the approximation of multivariate functions. From a more general formulation of the variational problem of regularization they derive the following approximation scheme, instead of equation (1):

$$f^*(\mathbf{x}) = \sum_{\alpha=1}^{n} c_\alpha G(\|\mathbf{x} - \mathbf{t}_\alpha)\|_W^2) + p(\mathbf{x}) \tag{4}$$

where the parameters $\mathbf{t}_\alpha$, which we call "centers," and the coefficients c_α are unknown, and are in general fewer than the data points ($n \leq N$). The term $p(\mathbf{x})$ is a polynomial that often can be neglected, though it may be useful to keep the constant and linear terms. The norm is a *weighted norm*

$$\|(\mathbf{x} - \mathbf{t}_\alpha)\|_W^2 = (\mathbf{x} - \mathbf{t}_\alpha)^T W^T W (\mathbf{x} - \mathbf{t}_\alpha) \tag{5}$$

where W is an unknown square matrix and the superscript T indicates the transpose operator. In the simple case of diagonal W the diagonal elements w_i assign a specific weight to each input coordinate, determining in fact the units of measure and the importance of each sensory input. In this

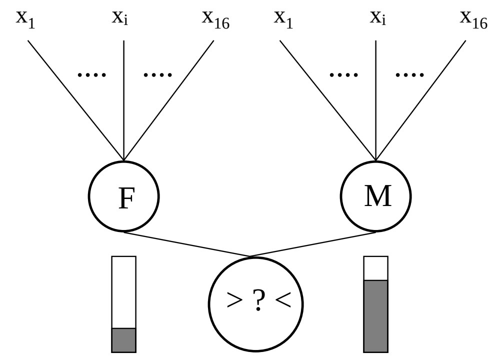


Fig. 2. The competing HyperBF Networks used for gender classification

formulation the learning stage is used to estimate not only the coefficients of the RBF expansion, but also the metric (problem dependent dimensionality reduction) and the position of the centers (optimal examples selection).

Let us think of a classification task in which the function range is represented by the closed interval $[0, 1]$. The value of the function can be interpreted as a *fuzzy predicate*. If a gaussian function is used the center of expansion is the only point at which the predicate assumes value 1: it can be effectively interpreted as a prototype (note that the use of HyperBF Networks for classification is directly related to Bayes estimation as pointed out in [10]).

Using a geometrical vector as input, gender classification has been attempted by using two competing networks: one for male recognition and one for female recognition (see Fig. 2 for the network structure). The gender to be associated to a given vector is taken to be that corresponding to the network with the greatest response. The reason for two competitive networks is that no threshold is then necessary. It is interesting to note how each of the networks is able to create a meaningful prototype of the class it represents. As can be seen in Figure4 the expansion center, which is a vector with components free to move during the "learning" process, has converged at the end of the training phase to what could be considered a caricature of a (fe)male face. It does not correspond to the average value on the separate subsets: it emphasizes the discriminating features. The learning stage is also able to change the metric to account for the different weight and significance of the different features. Of the sixteen features only three are given a noticeable weight: distance of eyebrow from eyes, eyebrows thickness and nose width. These are followed by the vertical position of nose and mouth and the two radii describing the lower chin shape; the remaining features are considered ineffective.

The database used for the classification experiments comprised 168 vectors equally distributed over 21 males and 21 females. Three different performances were measured:

- on the vectors of the training set (90% correct);
- on novel faces of people in the training set (86% correct);

- on faces of people not represented in the training set (79% correct)

The performance on a testing set, having null intersection with the training set, has been estimated with a *leave-one-out* strategy. Having n available examples, training was done on the first $n-1$ data leaving the last one for testing. The data set was then rotated, so that each of the available examples was used in turn as a testing example. The performance was estimated by taking the percentage of correct gender assessment on the resulting tests. The performance obtained is of 79% correct classifications.

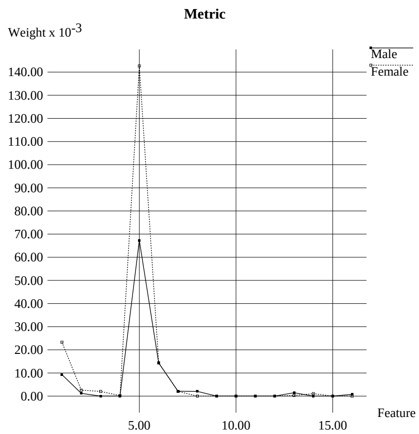


Fig. 3. Feature weights for gender classification as computed by the HyperBF Networks

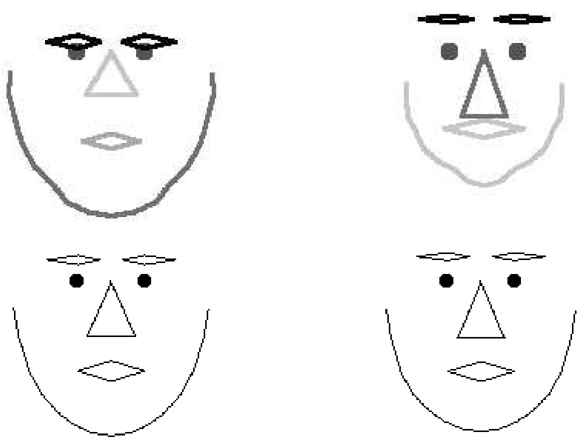


Fig. 4. TOP. The male prototype (left) and the female prototype (right) as synthesized by the HyperBF Networks with movable coefficients, centers and metric. The darker the feature, the more important it is according to the corresponding entries in the diagonal metric W. BOTTOM. The average male face (left) and the average female face (right)

Human performance in such classification tasks (as well as recognition) is widely believed to be nearly perfect. To assess the effective ability of people in gender classification we have performed some informal psycophysical experiments using as stimulation pattern a grey level image of the face from which the local average was subtracted (to make the different images as similar as possible). As Figure5 shows, no hair information was available (residual facial hair was masked out)

The database of stimuli was then presented one image

Fig. 5. Typical stimuli used in the experiments of human gender classification

after another on a computer screen and the subject was asked to press M for male and F for female without any time constraint. The results were surprinsing. An average score of 90% correct classification (on 17 subjects some of which familiar with a large subset of the people represented in the database). Classification performance was not impaired by the lack of familiarity with the database people. Informal chat with some of the subjects revealed that, at least consciously, eyebrow information was considered to be the most discriminating.

Note that no hair information has been used, both in our human and our computer gender classification experiments. This must be considered if these results are to be compared with other experiments reported in literature (see for example [6], where images included limited hair information).

3. Conclusion

Gender classification has been attempted using two competing HyperBF networks trained on a geometrical description of (fe)males faces. The resulting performance was of 79% correct classification (averaged on males and females) and must be confronted to a human performance of 90%. Analysis of the *internal representation* of the HyperBF networks shows that the networks have been able to effectively prototype (fe)male faces and that classification was achieved using a subset of the available features, similar to human strategies of gender classification.

References

[1] R. J. Baron. Mechanisms of human facial recognition. *International Journal of Man Machine Studies*, 15:137–178, 1981.

[2] H. Bourlard and Y. Kamp. Auto-association by multilayer perceptrons and singular value decomposition. *Biological Cybernetics*, 59(4/5):291–294, 1988.

[3] R. Brunelli and T. Poggio. Face Recognition: Features versus Templates. Technical Report 9110-04, I.R.S.T, 1991. to appear on IEEE Trans. on PAMI.

[4] G. Cottrell and M. Fleming. Face recognition using unsupervised feature extraction. In *Proceedings of the International Neural Network Conference*, 1990.

[5] I. Craw, H. Ellis, and J.R. Lishman. Automatic extraction of face features. *Pattern Recognition Letters*, 5:183–187, Feb. 87.

[6] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. Sexnet: A neural network identifies sex from human faces. In *Advances in Neural Information Processing Systems 3*, pages 572–577, 1991.

[7] T. Kanade. Picture processing by computer complex and recognition of human faces. Technical report, Kyoto University, Dept. of Information Science, 1973.

[8] M. Kirby and L. Sirovich. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, 1990.

[9] C. A. Micchelli. Interpolation of scattered data: Distance matrices and conditionally positive definite functions. *Constr. Approx*, 2:11–22, 1986.

[10] T. Poggio and F. Girosi. A theory of networks for approximation and learning. A.I. Memo No. 1140, Massachusetts Institute of Technology, 1989.

[11] T. Poggio and F. Girosi. Networks for Approximation and Learning. In *Proc. of the IEEE, Vol. 78*, pages 1481–1497, 1990.

[12] J. Sergent. Structural processing of faces. In A.W. Young and H.D. Ellis, editors, *Handbook of Research on Face Processing.* North-Holland, Amsterdam, 1989.

[13] M. Turk and A. Pentland. Eigenfaces for recognition. Technical Report 154, MIT Media Lab Vision and Modeling Group, 1990.