

An Investigation into the use of Partial-Faces for Face Recognition

Srinivas Gutta, Vasanth Philomin and Miroslav Trajković
Philips Research – USA
345 Scarborough Rd.
Briarcliff Manor, NY 10510
{Srinivas.Gutta, Vasath.Philomin, Miroslav.Trajkovic}@philips.com

Abstract

Even though numerous techniques for face recognition have been explored over the years, most research has primarily focussed on identification from full frontal/profile facial images. This paper conducts a first systemic study to assess the performance when using partial-faces for identification. Our specific approach considers an ensemble of Radial Basis Function (RBF) Networks. A specific advantage of using an ensemble is its ability to cope with the inherent variability in the image formation and data acquisition process. Our database consists of imagery corresponding to 150 unique subjects totaling to 3,000 facial images with $\pm 5^0$ rotation. Based on our experimental results, we observe that the average Cross Validation performance is the same even if only half the face image is used instead of the full-face image. Specifically we obtain 96 % when partial-faces are used and 97 % when full-faces are used.

1. Introduction

In order to interact socially, we must be able to process faces in a variety of ways. There is a vast amount of literature on social and cognitive psychology attesting to the impressive capabilities of humans at identifying familiar faces, as well as extracting information from both familiar and unfamiliar faces, including gender, race, and emotional state of the person [1]. Faces are accessible 'windows' into the mechanisms that govern our emotional and social lives. The face is a unique feature of human beings. Even the faces of "identical twins" differ in some respects. Humans can detect and identify faces in a scene with little or no effort even if only partial views of the faces are available. This skill is quite robust, despite large changes in the visual stimulus due to viewing

conditions, expression, aging, and distractions such as glasses or changes in hairstyle.

Automated recognition requires computer systems to look through many stored sets of characteristics ('the gallery') and pick the one that matches best those features of the unknown individual ('the probe'). In most practical scenarios there are two possible recognition tasks to be considered - (i) *Identification*: An image of an unknown individual is collected ('probe') and the identity is found searching a large set of images ('gallery'), and (ii) *Verification*: Rather than identifying a person, the system is now involved with verification and checks if a given probe belongs to a relatively small gallery, sometimes labeled as a set of intruders. In this paper we limit ourselves to the task of identification.

There are two major approaches for automated recognition of human faces. The first approach, the abstractive one, extracts (and measures) discrete *local* features 'indexes' for retrieving and identifying faces, so subsequently standard statistical pattern recognition techniques can be employed for probing amongst faces using these measurements. The other approach, the holistic one, conceptually related to template matching, attempts to recognize faces using *global* representations [2][3]. Common examples of these approaches include (a) Eigen Faces [4], (b) Elastic Bunch Graph Matching [5], (c) Linear Discriminant Analysis [6] and (d) Radial Basis Function Networks [7]. However, most research has primarily focussed on identification from full frontal/profile facial images. The only other paper that we are aware of that has performed identification from partial images is [8]. They have used partial face images – eye, nose and ear images for identification. They report an accuracy of 100 % recognition/rejection on a database of 720 images corresponding to 120 subjects.

In this paper we describe a face recognition system based on Radial Basis Function Networks. In our experiments we limit our attention to identification of faces from approximately frontal images. Specifically, we attempt to identify subjects from partial face images. As an example if the face image is of dimension 64x72, our input to the recognition module is of dimension 32x72. The paper is organized as follows. The overall architecture is described in Section 2, while the system and the tools developed are described in Section 3. Extensive experiments were carried out to validate the system and they are described in Section 4. The paper concludes in Section 5 with an assessment of our results and some mention of future work.

2. Face Recognition

An overall architecture, appropriate for face recognition, is shown in Fig. 1. Face recognition usually starts with the detection of a pattern as a face, proceeds by normalizing the face image to account for geometrical and illumination changes using information from the box surrounding the face and/or eye locations, and finally it identifies the face using appropriate image representation and classification algorithms. The tools needed to detect face patterns and normalize them are discussed elsewhere [9], while this paper describes only the tools developed to realize and implement those stages of face recognition involved in identification tasks. The specific task under consideration is the identification of subjects from partial face images on a database of 150 subjects comprising of approximately 3,000 images.

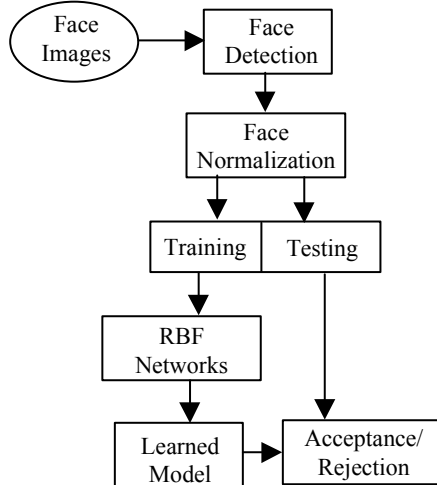


Figure 1 Automated Face Recognition Architecture

3. Radial Basis Function Networks

The construction of the RBF network involves three different layers. The input layer is made up of source nodes (sensory units). The second layer is a hidden layer whose goal is to cluster the data and reduce its dimensionality. The output layer supplies the response of the network to the activation patterns applied to the input layer. The transformation from the input space to the hidden-unit space is *non-linear*, whereas the transformation from the hidden-unit space to the output space is *linear*. In particular, we note that a RBF classifier can be viewed in two ways [10]. One is to interpret the RBF classifier as a set of kernel functions that expand input vectors into a high-dimensional space, trying to take advantage of the mathematical fact that a classification problem cast into a high-dimensional space is more likely to be linearly separable than one in a low-dimensional space. Another view is to interpret the RBF classifier as a function-mapping interpolation method that tries to construct hypersurfaces, one for each class, by taking a linear combination of the Basis Functions (BF). These hypersurfaces can be viewed as discriminant functions, where the surface has a high value for the class it represents and a low value for all others. An unknown input vector is classified as belonging to the class associated with the hypersurface with the largest output at that point. In this case the BFs do not serve as a basis for a high-dimensional space, but as components in a finite expansion of the desired hypersurface where the component coefficients, (the weights) have to be trained.

An RBF classifier has architecture very similar to that of a traditional three-layer back-propagation network. Connections between the input and middle layers have unit weights and, as a result, do not have to be trained. Nodes in the middle layer, called BF nodes, produce a localized response to the input using Gaussian kernels. Each hidden unit can be viewed as a localized receptive field (RF). The hidden layer is trained using k-means clustering. The most common basis function (BF) used are Gaussians, where the activation level y_i of the hidden unit i is given by:

$$y_i = \Phi_i(\|X - \mu_i\|) = \exp\left[-\sum_{k=1}^D \frac{(x_k - \mu_{ik})^2}{2h\sigma_{ik}^2 o}\right]$$

where h is a proportionality constant for the variance, x_k is the k th component of the input vector $X=[x_1, x_2, \dots, x_D]$, and μ_{ik} and σ_{ik}^2 are the k th components of the mean and variance vectors, and o is the overlap factor,

respectively, of basis function node i . The outputs of the hidden unit lie between 0 and 1, and could be possibly interpreted as fuzzy memberships; the closer the input to the center of the Gaussian, the larger the response of the node. The activation level Z_j of an output unit is given by:

$$Z_j = \sum_i w_{ij} y_i + w_{0j}$$

where Z_j is the output of the j th output node, y_i is the activation of the i th BF node, w_{ij} is the weight connecting the i th BF node to the j th output node, and w_{0j} is the bias or the threshold of the j th output node. The bias comes from the weights associated with a BF node that has a constant unit output regardless of the input. An unknown vector X is classified as belonging to the class associated with the output node j with the largest output Z_j .

The RBF input consists of n normalized face images pixels fed to the network as 1D vectors. The hidden (unsupervised) layer, implements an enhanced k-means clustering procedure, where both the number of Gaussian cluster nodes and their variance are dynamically set. The number of clusters varies, in steps of 5, from 1/5 of the number of training images to n , the total number of training images. The width of the Gaussian for each cluster, is set to the maximum of *{the distance between the center of the cluster and the member of the cluster that is farthest away - within class diameter, the distance between the center of the cluster and closest pattern from all other clusters}* multiplied by an overlap factor o , in our experiment equal to 2. The width is further dynamically refined using different proportionality constants h . The hidden layer yields the equivalent of a functional facial base, where each cluster node encodes some common characteristics across the face space. The output (supervised) layer maps face encodings ('expansions') along such a space to their corresponding class and finds the corresponding expansion ('weight') coefficients using pseudoinverse techniques. In our case the number of nodes in the output layer correspond to the number of people we wish to identify. As an example, if we wish to identify 50 subjects then the output layer would have 50 nodes. The number of nodes in the input layer on the other hand corresponds to the size of the input image. Note that the number of clusters is frozen for that configuration (number of clusters and specific proportionality constant h) which yields 100 % accuracy when tested on the same training images.

For a connectionist architecture to be successful it has to cope with the variability available in the data acquisition process. One possible solution to the above problem is to implement the equivalent of query by consensus using ensembles of radial basis functions (ERBF). Ensembles are defined in terms of their specific topology (connections and RBF nodes) and the data they are trained on. Specifically, both original data and distortions caused by geometrical changes and blur are used to induce robustness to those very distortions via generalization [7].

The ERBF architecture is shown in Fig. 2. Each RBF component is further defined in terms of three RBF nodes, each of which specified in terms of the number of clusters and the overlap factors. The overlap factors o , defined earlier, for the RBF nodes RBF (11, 21, 31), RBF(12, 22, 32), and RBF(13, 23, 33) are set to 2, 2.5, and 3, respectively. The same RBF nodes were trained on original images, and on the same original images with either some Gaussian noise added or subject to some degree of geometrical ('rotation'), respectively. The intermediate nodes C_1 , C_2 , and C_3 act as buffers for the transfer of the normalized images to the various RBF components. Training is performed until 100% recognition accuracy is achieved for each RBF node. The nine output vectors generated by the RBF nodes are passed to a *judge* who would make a decision on whether the probe ('input') belongs to that particular class or not. The specific decision used is - if the average of 5 of the 9 network outputs is greater than θ then that probe belongs to that class.

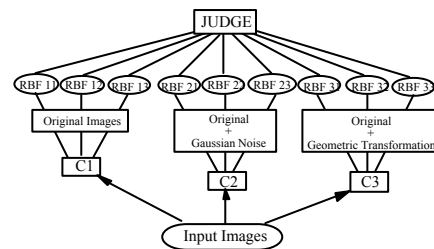


Figure 2 ERBF1 Architecture

4. Experiments

The number of unique individuals in our database corresponds to 150 subjects. During image acquisition each subject was asked to sit in front of the computer equipped with a Matrox frame grabber and a Philips CCD camera. The distance from the subject and the camera was approximately 3 feet. Top view geometry for facial database acquisition is shown below in Fig. 3.

Each subject was asked to first look at the camera for approximately 5 seconds and turn his/her head $\pm 5^\circ$. The subjects were asked to make different kind of facial expressions which include smiling, surprise, etc. The frame grabber was set to acquire imagery at the rate of 5 fps. The images were acquired at a resolution of 640x480 pixels and encoded in 255 gray scale levels. The images are then passed to the face detection module [9]. Detected faces greater than the set threshold of 0.95 were stored. The faces are then normalized to account for geometrical and illumination changes using information about the eye location. The final face obtained at the end of detection and normalization is of a standard resolution of 64x72 pixels. Since we know the location of the eyes from the detection module, we create the partial-face by cutting the normalized facial image vertically at the point where the distance from one end of the image to the center is 1/2 the eye distance. A sample set of face images and their corresponding partial faces are shown in Figs. 4 and 5 respectively.

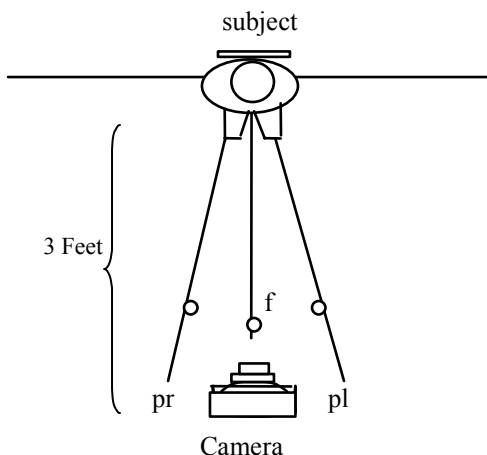


Figure 3 Top View Geometry of Facial Database Acquisition. 'f': Frontal, 'pr': Profile Right, 'pl': Profile Left.



Figure 4 Examples of Detected Frontal Faces



Figure 5 Examples of Partial Faces

In Section 4.1 we report on the experiments conducted for the identification of subjects from partial faces, while in Section 4.2 we report the results for the full-face recognition task.

4.1 Partial Face Recognition

First we report on experiments when only one instance of the subjects facial image was used for training followed by the case in which multiple instances were used for training. The training and testing strategy used in both cases is similar to that of k - fold cross validation (CV) [11]. In k - fold cross validation, the cases are randomly divided into k mutually exclusive partitions of approximately equal size. The cases not found in one partition are used for training, and the resulting classifier is then tested on the partition left out of training. The average error rates over all k partitions are the CV error rate. Each partition in our case consists of images corresponding to 50 unique subjects. Each CV cycle consists of 20 iterations. As an example, the first CV cycle on its first iteration would randomly pick one image corresponding to each of the 50 subjects ('gallery') in the first partition while testing on the remaining images of the subjects in the gallery plus the images ('probes') corresponding to 100 subjects. For each cycle this process is repeated for a total of 20 times. Table 1 shows the average CV results when the threshold was set at 0.70.

CV Cycle	Accepted (Correct) %	False Negative %	Accepted (Correct) %	False Negative %
A	75	25	80	20
B	80	20	78	22
C	78	22	80	20
Avg. Cycle	77.67	22.33	79.33	20.67

Table 1 Average CV Results for Partial-Face Recognition when Gallery is 1 Image per Individual

CV Cycle	Accepted (Correct) %	False Negative %	Accepted (Correct) %	False Negative %
A	85	12	92	8
B	83	10	91	9
C	86	9	94	6
Avg. Cycle	84.67	15.33	92.33	7.67

Table 2 Average CV Results for Partial-Face Recognition when Gallery is 5 Images per Individual

Tables 2 and 3 show the average CV results when the number of gallery images was increased to 5 and 9 respectively. Table 4 shows the average CV results across all cycles and iterations when an ensemble is used. The specific procedure used for training and testing each of the nine networks remains the same as that used for a single RBF network.

CV Cycle	Accepted (Correct) %	False Negative %	Accepted (Correct) %	False Negative %
A	90	10	96	4
B	92	8	95	5
C	92	8	95	5
Avg. Cycle	91.33	8.67	95.33	4.67

Table 3 Average CV Results for Partial-Face Recognition when Gallery is 9 Images per Individual

CV Cycle	Accepted (Correct) %	False Negative %	Accepted (Correct) %	False Negative %
G1	85	15	87	13
G5	91	9	93	7
G9	96	4	97	3

Table 4 Average CV Results for Partial-Face Recognition when Gallery is 1, 5 and 9 Images per Individual

4.2 Full Face Recognition

We also experimented to assess the performance when full faces were used. In the case of full faces, the images obtained after they were normalized were

passed directly to the RBF network for training/testing. The specific training and testing procedure remains the same as that used in section 4.1. The average CV results are shown in Tables 5, 6 and 7 respectively, while Table 8 shows the results when an ensemble is used on full faces. We also ran further experiments to assess the performance change as the amount of information present in the image was increased from 50 % to 100 % in steps of 10 %. Figure 6 below graphs the result when an ensemble was used.

CV Cycle	Accepted (Correct) %	False Negative %	Accepted (Correct) %	False Negative %
A	86	14	92	8
B	88	12	90	10
C	92	8	91	9
Avg. Cycle	88.67	11.33	91	9

Table 5 Average CV Results for Full-Face Recognition when Gallery is 1 Image per Individual

CV Cycle	Accepted (Correct) %	False Negative %	Accepted (Correct) %	False Negative %
A	94	10	94	6
B	88	12	96	4
C	95	5	95	5
Avg. Cycle	92.33	7.67	95	5

Table 6 Average CV Results for Full-Face Recognition when Gallery is 5 Images per Individual

CV Cycle	Accepted (Correct) %	False Negative %	Accepted (Correct) %	False Negative %
A	96	4	98	2
B	98	2	95	5
C	94	6	95	5
Avg. Cycle	96	4	96	4

Table 7 Average CV Results for Full-Face Recognition when Gallery is 9 Images per Individual

4. Conclusions

We have proposed in this paper an ensemble of RBF networks for partial-face identification and showed their feasibility on a collection of 3,000 face images corresponding to 150 subjects with $\pm 5^\circ$ rotation. Cross Validation (CV) results yield an average accuracy rate of (a) 96% on the partial-face identification task and (b) 97% on the full-face identification task. Based on the experimental results, we believe that a partial-face is sufficient for accurate identification. Moreover, using partial faces will also reduce the amount of computational power and storage requirements by a significant amount.

We are currently adapting the RBF network so that it could be trained on a full-face image but accept a partial-face image as the probe during testing. This will allow existing systems to do face recognition from partial images without having to be retrained.

CV Cycle	Accepted (Correct) %	False Negative %	Accepted (Correct) %	False Negative %
G1	92	8	94	6
G5	95	5	96	4
G9	97	3	99	1

Table 8 Average CV Results for Full-Face Recognition when Gallery is 1, 5 and 9 Images per Individual

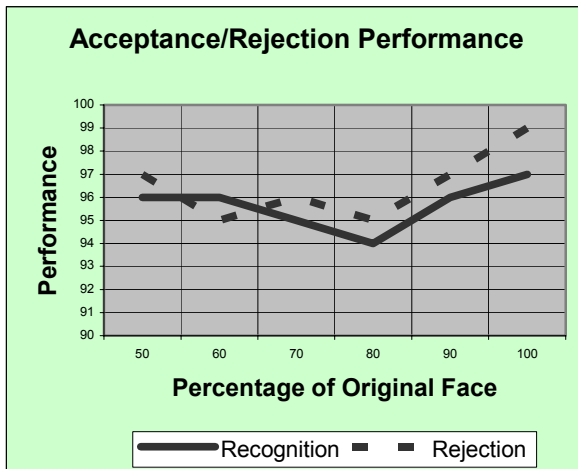


Figure 6 Average Performance vs. Amount of Image Information Used for an Ensemble

REFERENCES

- [1] D. Valentin, H. Abdi, A. Toole and G.W. Cottrell, Connectionist Models of Face Processing: A Survey, *Pattern Recognition* 27(9): 1209-1230, 1994.
- [2] H. Wechsler, P.J. Phillips, V. Bruce, F. F. Soulie and T. S. Huang, *Face Recognition: From Theory to Applications*, Springer-Verlag, New York, 1998.
- [3] S. Gong, S. J. McKenna and A. Psarrou, *Dynamic Vision: From Images to Face Recognition*, Imperial College Press, London, 2000.
- [4] M. Turk and A. Pentland, Eigenfaces for Recognition, *J. Cognitive Neuroscience* 3, 71-86.
- [5] L. Wiskott, J. M. Fellous, N. Krüger and C. Malsburg, Face Recognition by Elastic Graph Matching, *IEEE PAMI* 19(7): 775-779, 1996.
- [6] K. Etemad and R. Chellappa, Discriminant Analysis for Recognition of Human Face Images, *J. Optical Society of America* 14: 1724-1733, 1997.
- [7] S. Gutta and H. Wechsler, Face Recognition using Hybrid Classifiers. *Pattern Recognition* 30(4): 539-553, 1997.
- [8] K. Sato, S. Shah and J. K. Aggarwal, Partial Face Recognition using Radial Basis Function networks, in *Proc. of the 3rd International Conference on face and Gesture Recognition* 288-293, Nara, Japan, 1998.
- [9] A. Colmenarez, B. Frey and T. S. Huang, Detection and Tracking of Faces and Facial Features, in *Proc. of International Conference on Image Processing*, Kobe, Japan, 1999.
- [10] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd Edition, Prentice Hall, New Jersey, 1999.
- [11] S. M. Weiss and C. A. Kulikowski, *Computer Systems That Learn*, Morgan Kaufmann, Palo Alto, 1991.