FAST RECOGNITION OF MULTIPLE FACES USING MCM [1]

Manoj Seshadrinathan and Jezekiel Ben-Arie

ECE Dept., M/C 154, University Of Illinois at Chicago,
851 S.Morgan St., Chicago, IL 60607, USA
Phone/Fax:001-312-996-2648, E-mail:benarie@ece.uic.edu

Abstract: In this paper we present the novel Minimal Classification Method (MCM) that provides an elegant and computationally efficient approach to the problem of multiple face recognition. In the classification we use a special set of filters which yield low values of correlation with a given set of images and large values with other natural images. Such filters are called anti-faces. The Minimal Classification Method can efficiently recognize individual images from a large set of 2^k images using only k filter sets whereas other methods require much more classifiers. We use here anti-faces, but the same approach can be implemented with other pattern recognition methods such as SVM or Eigenfaces.

Keywords: Minimal Classification Method(MCM), SVM, Rayleigh-Ritz theorem, face recognition, smoothness

1. INTRODUCTION

Our Minimal Classification Method (MCM) is a novel method that provides an computationally efficient approach to pattern recognition problems such as face recognition. Our method works with various recognition schemes that are capable of classifying multiple patterns. For instance, methods like the Support Vector Machines [4], Eigenfaces [7] and Anti-faces [1] are methods that can classify multiple patterns (templates). Though the schemes by themselves are robust and efficient in discriminating between pattern classes, the actual process of recognition of individual patterns requires a series of pattern class discriminations (i.e classifications). For example, in [2], the tournament knock-out method is employed where 2^k classes require $2^k - 1$ comparisons to identify an individual pattern.

The problem that we address can be stated as: *Given a database of n patterns and a pattern recognition approach capable of classifying multiple pattens, find a method that uses a minimal number of pattern classification operations to identify uniquely any pattern in the set.* For the above problem, using our Minimal Classification approach, only k classifiers are necessary instead of $2^k - 1$ in the tournament method. This results in significant computational savings.

In this paper, we apply the Minimal Classification Method to the problem of multiple face Recognition. For the multiple template classification scheme, we employ novel filters called Anti-Faces[1]. These filters yield low values of inner product with the required class and high values with any other natural image. Such filters have been recently used for single face recognition[1]. However, no effort was made to recognize individual faces in a multiple face database with this method. The anti-face filters have to yield small values of inner product with the required images and large values with all other images. Hence, the filters have to satisfy the following conditions 1) Yield low values of inner product with the required image set. 2) Have to be smooth so that they yield high value of inner product with other natural images. Our method for finding these filters uses the Rayleigh-Ritz theorem.

[1] This work was supported by the National Science Foundation (NSF) Grants No. IIS-0208300 and IIS-9979774.

2. MULTIPLE TEMPLATE RECOGNITION

In this paper, we focus on the Minimal Classification Method(MCM) that efficiently groups patterns so that they can be classified with minimal number of pattern classification operations. We use the Anti-faces filters for the recognition phase since they have been proved as a novel and efficient method for face recognition [1]. The filters are found based on the Rayleigh Ritz theorem described in the next subsection.

The Rayleigh Ritz Theorem: (Characterization of Eigenvalues) is stated as follows: If one has to maximize/minimize a quadratic form $\underline{\mathbf{X}}^T A \underline{\mathbf{X}}$ subject to the constraint $\underline{\mathbf{X}}^T \underline{\mathbf{X}} = 1$ then the largest eigenvalue of A achieves the maximum and the smallest eigenvalue of A achieves the minimum and the solution $\underline{\mathbf{X}}$ is the corresponding eigenvector. From this theorem it follows that the minimal value of the quadratic form $\underline{\mathbf{X}}^T A \underline{\mathbf{X}}$ is obtained when $\underline{\mathbf{X}}$ is equal to the eigenvector that corresponds to the smallest eigenvalue λ_1. Also, this minimal value is equal to λ_1. We apply the Rayleigh Ritz theorem to find the filters. Anti-Faces as mentioned before are novel filters that give low value of the dot product with the required image and high values for any other image. Multiple face recognition involves recognition of all the faces in a given model database. This can be posed as a multiple template recognition problem. Given a set of images we need to recognize instances of all the model images in test images. This implies that each filter gives a low value of dot product with the set of model faces (the model class) and high value with other images. Given a set of images $Q = \{\underline{t}_i\}$, we arrange them as columns of a matrix A. Next, we find the filter $\underline{\mathbf{X}}$ which minimizes the sum of the squares of the inner products of $\underline{\mathbf{X}}$ with all the images $\underline{\mathbf{t}} \in Q$, that is

$$\sum_{t \epsilon Q} (\underline{\mathbf{X}}, \underline{\mathbf{t}})^2 = \underline{\mathbf{X}}^T A A^T \underline{\mathbf{X}} \longrightarrow min. \quad (1)$$

where A is the matrix of image vectors as its columns. The filter $\underline{\mathbf{X}}$ is normalized to a unit norm and zero mean (sum of filter components). **Smoothness Constraint:** Finding a vector $\underline{\mathbf{X}}$ that minimizes the sum of squares of the filter with the images is not sufficient. There are many vectors $\underline{\mathbf{X}}$ that yield low inner product with all the corresponding faces, but they also may have small values of inner products with images that are not members of the training set. Statistically, most natural images are smooth and the expectation of the absolute value of the inner product of two smooth vectors is large [1]. Hence, it follows that the filter $\underline{\mathbf{X}}$ has to be smooth in order to yield high values of inner product with random natural images. In our case, the filter $\underline{\mathbf{X}}$ has also to be smooth in order to give high correlation with the images other than faces. The Boltzman distribution has been proved to be a proper model for natural images[3]. It assigns to any image $\underline{\mathbf{F}}$ a probability proportional to the exponent of the negative of a roughness measure for the particular image. Thus, $P(\underline{\mathbf{F}})\ \alpha\ exp(-S(\underline{\mathbf{F}}))$, where $S(\underline{\mathbf{F}})$ is a roughness measure of the image expressed by the coefficients of the Discrete Cosine Transform (DCT).Representing an image $\underline{\mathbf{F}}$ by the Discrete cosine transform(DCT), one can obtain a measure of the roughness $S(\underline{\mathbf{F}})$ given by the expression:

$$S(\underline{\mathbf{F}}) = (k^2 + l^2) f^2(k, l) \quad (2)$$

where k and l are the vertical and horizontal frequencies of the image and $f(k, l)$ are its DCT coefficients. To prove that the absolute value of dot product of two random images (sampled from the probability space) is usually large, one has the calculate the expectation of the square of inner product of an image with a random image. For a given image $\underline{F_1}$ the expectation of the inner product with a random image $\underline{F_2}$ can be derived[1] as

$$E[(\underline{F_1}, \underline{F_2})^2] = \sum_{(k,l)(0,0)} \frac{f_1^2(k, l)}{(k^2 + l^2)^{\frac{3}{2}}} \quad (3)$$

where $f_1(k,l)$ are the DCT coefficients of the image F_1. Hence, it is obvious that for $E[(\underline{F_1}, \underline{F_2})^2]$ to be large, F_1 has to be smooth since then the dominant values of the DCT would be concentrated in the smaller frequencies k,l which have smaller divisors. As mentioned, the filter $\underline{\mathbf{X}}$ should be smooth in order to yield high values of dot product with random natural images. The roughness of the filter $\mathbf{S(\underline{X})}$ is given by equation (5) as

$$S(\underline{X}) = (k^2 + l^2)\underline{\mathbf{Y}}^2(k,l) \tag{4}$$

where $\underline{\mathbf{Y}}(k,l)$ are the DCT coefficients of the filter $\underline{\mathbf{X}}$. $S(\underline{\mathbf{X}})$ can be written as $\underline{\mathbf{Y}}^T B \underline{\mathbf{Y}}$ where matrix B is a diagonal matrix of $(k^2 + l^2)$ values of the filter and $\underline{\mathbf{Y}}$ is a vector composed of the DCT coefficients of $\underline{\mathbf{X}}$ i.e $\underline{\mathbf{Y}}(k,l)$. The computation of the filters are carried out in the frequency domain since the quadratic expression of the roughness $\underline{\mathbf{Y}}^T B \underline{\mathbf{Y}}$ is more simplified and the matrix B is diagonal. Note that the images are normalized to zero mean and a unit norm. Here, to accommodate for minimum roughness the equation (1) is modified as:

$$\sum_{t \epsilon Q} (\underline{\mathbf{Y}}, \underline{\mathbf{t}})^2 + \lambda S(\underline{\mathbf{Y}}) \rightarrow \mathbf{min} \tag{5}$$

where $S(\underline{\mathbf{Y}})$ is the roughness of the filter $\underline{\mathbf{Y}}$. Using λ as a parameter the expression to be minimized is now:

$$\underline{\mathbf{Y}}(DD^T + \lambda B)\underline{\mathbf{Y}}^T \rightarrow \mathbf{min} \tag{6}$$

where D is the matrix of the DCT coefficients of the images where each column represents an image and B matrix is the diagonal matrix of the $(k^2 + l^2)$ values of the filter. The desired filter is the inverse DCT transform of $\underline{\mathbf{Y}}$. With $\lambda = 0$ we get the minimized sum of square of inner products. However this solution is not a smooth filter. So, as the value of λ increases the filter becomes more smooth probably at the cost of not giving the least dot product. The value of λ can be found by a binary search and be adjusted such that the value of $\sum_{t \epsilon Q} (\underline{\mathbf{X}}, \underline{\mathbf{t}})^2$ is below a desired value δ. The $\underline{\mathbf{X}}$ that minimizes this equation can be found by the Rayleigh-Ritz theorem.

3. THE MINIMAL CLASSIFICATION METHOD(MCM)

Our Minimal Classification Method minimizes the number of classifications required to classify a pattern in a pattern set. It can be applied to any classification problem and with any multiple template recognition scheme. We just need k pattern recognition operations for classifying 2^k patters. This is minimal since each pattern in a database of 2^k patterns can be represented by no fewer than k bits. Our method has an advantage over the prevalent tournament method [2] proposed in [2] and in other works. In the tournament method each and every image in the training set are considered as players. If there are 2^k players then in the first round by comparison 2^{k-1} players are eliminated and the winners proceed to the second round where similar elimination takes place. Thus, the tournament method for 2^k images requires $2^k - 1$ pattern classification operations in order to recognize an image uniquely. **In contrast to the tournament approach, we need just a maximum of k filters for identifying 2^k images.** For example, with 256 images the tournament method requires 255 comparisons whereas we need only 8. This saves a lot of computations.

Method: Our Minimal Classification Method minimizes the number of comparisons required to classify a pattern in a pattern set. It can be applied to any classification problem. Each

[2] The tournament method is elaborated more in the Appendix

pattern in a database is assigned an integer code number, g. If the total number of patterns is 2^k, k bits are needed to represent each integer g in a binary form. Next, the set of 2^k images is divided into two subsets in k different divisions, where each division corresponds to a bit in the integer g. For example, the first division is according to the least significant bit in g and divides the images into 2 subsets one of even g and the other with odd g. Our classification method is minimal since the number of comparisons required can not be less than k if one wants to classify each pattern in a 2^k pattern set. In any database of images each of the images are assigned an integer code number g expressed by a k bit binary number. We start by finding filters, using the method in Section 2, for the subset of the images which have the first bit of their bit code as 1. Subsequently, we follow the same procedure for finding the set of filters for each of the bits in g. Hence, we obtain k filters.

Recognition: To recognize whether an arbitrary face belongs to the database we run the filters sequentially over the image and we set the bit corresponding to each filter as one if the inner product is below a certain threshold (10^{-4}) and zero if the inner product is higher. In this way each filter recognizes whether the image belongs to the particular subset or not. We get a resultant binary code that corresponds to the image which is recognized and if the image does not belong to any of the subsets the image code is $00\cdots0$. Our Minimal Classification Method (MCM) works with any pattern set and not just images. The Anti-faces were used since they proved to be good for multiple template recognition. Any other method, such as the SVM [4], which recognizes multiple templates can also be used with the MCM for computationally efficient recognition of multiple faces. The Anti-faces were chosen since results from [1] show that it has significant advantages over methods like the Fisher Linear Discriminant.

4. CONCLUSION AND EXPERIMENTAL RESULTS

Results with the Anti-faces: To apply this problem of Multiple Face Recognition by Minimal Classification we created a face model set of 126 different human faces. The model database was created using 40 images from the ORL face database[6] and 86 images from the CVL face database[5]. All images were adjusted to size 56 x 46. All the images are also normalized to zero mean and a norm of 1.

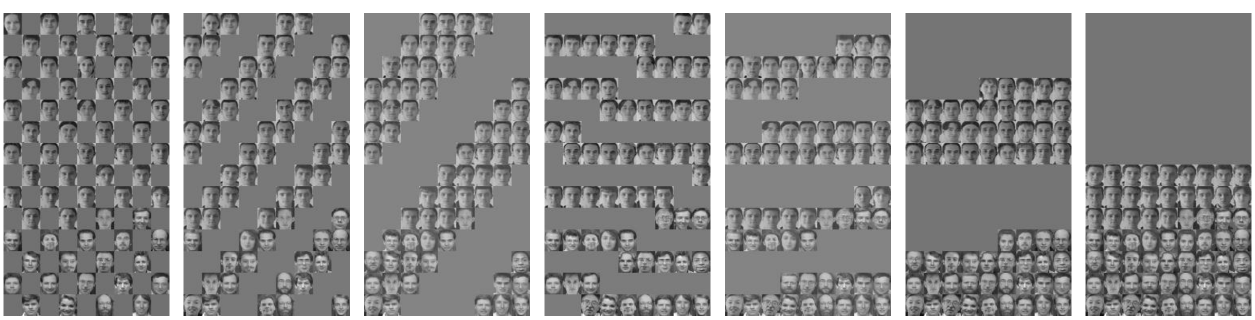

Fig. 1 The groups corresponding to all the 7 bits where for each image one of the bits is set to 1 (from LSB to MSB)

Each image is represented by a seven bit binary code number. We group the images into eight subsets based on the value of each of their bits. The filters are found for the different subsets from equation (5) using the Rayleigh-Ritz theorem and incorporating the smoothness criterion. In the recognition stage, the filters are correlated with all the test images

Fig. 2 a.) Random database b.) The 20 faces belonging to the original database correctly retrieved from a Random Image set c.) and d.) Examples of the Anti-face filters.

sequentially and the 7 bit code is found. This code corresponds to the identity number of the image recognized. The output for all other faces which did not belong to the database is 0. Totally, we can recognize up to 126 faces with just seven filters. In our experiments, any image out of the 126 images in our database can be uniquely recognized using these filters. In order to test the performance further, some of the images were subjected to rotations and shifts. It is to be noted that as the smoothness increases the anti-faces become more tolerant to noise and other perturbations in the image but also may give rise to spurious false alarms. In our case, when we use a threshold of 10^{-4}, the detection was accurate up to 6 degrees of rotation. The recognition was 100% accurate for above Signal to Noise Ratios (SNR) of 30dB. It falls rapidly below SNRs of 17dB. A plot of Recognition Rate vs. SNR is shown in Figure 3 (left). The method is also tested to retrieve images belonging to the database from an arbitrary collection of images. The filters are run over all the images in the collection. And all images which correspond to the same code are retrieved as similar images. To test the capability to recognize similar images, twenty images from the ORL face set are dispersed in a database of 40 random images (not in the original training data set). All the images were accurately recognized and retrieved. The images are described in Figure 2 ((a) and (b)). We see that all the 20 images are correctly retrieved and we also obtain their position (bit code) in the original database. All the other images which do not belong to the database receive a code $00 \cdots 0$.

Comparison with the tournament method: As mentioned before, our method can be used with SVM[4] and other multiple pattern recognition methods. If we have two classes P_1 and P_2 consisting of a set of patterns the support vector machine finds the hyperplane that maximizes the margin between the two classes P_1 and P_2 and minimizes the misclassification errors. Given fixed probability distributions, this hyperplane that minimizes the misclassification errors not only of the training set but also of the test set (which are unknown), is called the Optimal Separating Hyperplane (OSH)[4]. In [8] SVMs have been used successfully to classify faces from non-faces. The tournament method proposed in [2] finds the OSH between each pair i and j in the training set. Thus, they obtain OSH between each pair. When an image from the test set is to be recognized they follow the tournament method in which each object is a player. The test image is classified according to the OSH related to the two images involved. If there are 2^k patterns then 2^{k-1} classifications are made and 2^{k-1} patterns are eliminated. The remaining patterns advance to the next round. Hence, a total of $2^k - 1$ classifications are required. In our MCM we assign integer codes to the images. Each OSH separates the even bit group from the odd group. Hence, for k bits we have k OSH. Now it is enough if we make just k classifications to uniquely recognize an pattern from the 2^k patterns. This is best illustrated by the following example.

To summarize, as opposed to the tournament method that requires $2^k - 1$ OSH in a database of 2^k images, the MCM requires only k OSH and one can uniquely identify each image using just these k classifications. Hence, we obtain a great reduction in complexity and an increase

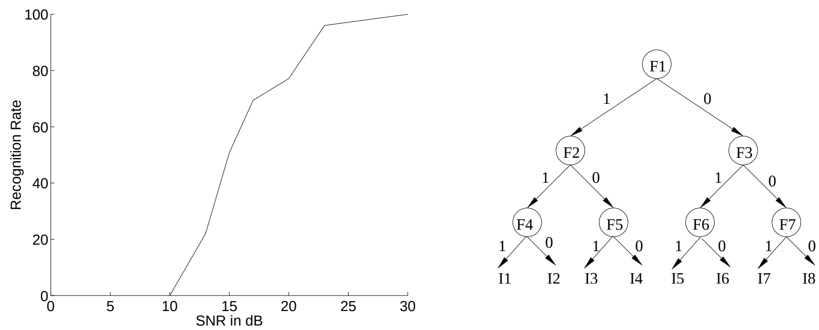

Fig. 3 Plot of Recognition Rate vs. SNR (left) and Binary Tree Approach for recognition of 8 faces (right)

in efficiency through the use of the Minimal Classification Method.

Application with Binary Tree: Another approach for fast and efficient recognition of multiple faces would be the Binary Tree approach. If the number of patterns are 2^k , then the filter (classifier) at the top node of the tree classifies 2^{k-1} patterns from the other 2^{k-1} patterns. Each node has two subnodes that further classify the subsets into two groups. Figure 3 (right) illustrates the Binary Tree approach for recognition of 8 faces. We see that, using the Binary Tree approach, any pattern in the set of 2^k patterns can be uniquely identified using just k comparisons. But the number of classifiers needed is 2^{k-1}. Hence, this method is not minimal with respect to the number of classifiers. Based on the number and complexity of the patterns and the multiple template recognition scheme used, either the Binary Tree approach or the Minimal Classification approach can be used for efficient recognition of multiple faces.

REFERENCES

[1] Keren.D,Osadchy.M and Gotsman.C, "Anti-faces: A novel fast method for image detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 23, pp. 747-761, July 2001.

[2] Massimiliano Pontil and Alessandro Verri, "Support Vector Machines for 3D Object Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence Vol 20, pp. 637–646, June 1998.

[3] Daniel Keren and Michael Werman "Probabilistic Analysis of Regularization," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 15, No. 10, pp. 982-995, 1993.

[4] V. N. Vapnik, The Nature of Statistical Learning Theory, Berlin: Springer-Verlag 1995.

[5] CVL Face Database obtained from Computer Vision Laboratory, Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia and SCV, PTERS, Velenje, http://www.lrv.fri.uni-lj.si/facedb.html

[6] ORL Face Database obtained from AT&T Labs, Cambridge.

[7] Matthew A. Turk, Alex P. Pentland, "Face Recognition using Eigenfaces," Intl. Conference on Computer Vision and Pattern Recognition, pp 586-591, 1991.

[8] Edgar Osuna, Robert Freund, Federico Girosi, "Training Support Vector Machines: An Application to Face Detection," Proceedings of CVPR, pp. 130-136, 1997