# Probabilistic Face Recognition using Bayesian Analysis

**Nikhil Dandekar**
Department of Computer Science
Rutgers University

## Abstract

The project aims at implementing a face recognition system based on Bayesian analysis of difference images. It converts the problem of face recognition into a two-class clustering problem, which then can be conveniently solved using Bayesian decision theory. The advantages of this approach are - better performance than conventional approaches and increased computational efficiency especially when dealing with very large databases.

## 1 Introduction

Face recognition is a well-researched Computer Vision problem from as far back as the 1960s. Early face recognition systems relied on the geometry of *fiducial* points such as eye/nose/mouth corners and their spatial relationships also known as the "feature-based" paradigm. In the late 1980s, researchers began using appearance or texture of facial images often as raw 2D inputs to their systems in what is known as the "template-based" paradigm. A study by Brunelli & Poggio [6], compares the two approaches. The current state-of-the-art in face recognition is characterized by a family of subspace methods originated by Turk & Pentland's "eigenfaces" [7] approach. A number of extensions and variants of this technique have been proposed since then.

A significant departure from the previous work is the research done by Moghaddam and Pentland [1, 3, 4]. Their *probabilistic* classifiers differ from traditional classifiers in two important ways. First, their system uses a Bayesian i.e. a *Maximum A Posteriori* classifier. Second, they cast the multi-class problem of distinguishing among images of different subjects into a problem of distinguishing between *intrapersonal* and *extrapersonal* difference images. This project aims at using this approach and applying it to solve the face recognition problem.

## 2 Concept

The project aims at implementing a face recognition system based on a probabilistic measure of similarity between faces. The main idea has been proposed in [1], and more explanation of the concepts involved in [1] can be found in [3] & [4]. The main concept involved is that image intensity differences i.e. the difference image obtained by subtracting the intensities of 2 images can be classified into either of the 2 classes - intrapersonal variations (corresponding to different facial expressions of the same person) and extrapersonal variations (corresponding to variations between different individuals). Thus if the difference image lies in the class of intrapersonal variations, we can say that the two images which formed the difference image are of the same individual. Likewise, if the difference image lies in the class of extrapersonal variations, we can say that the two images that formed the difference image are of different individuals.

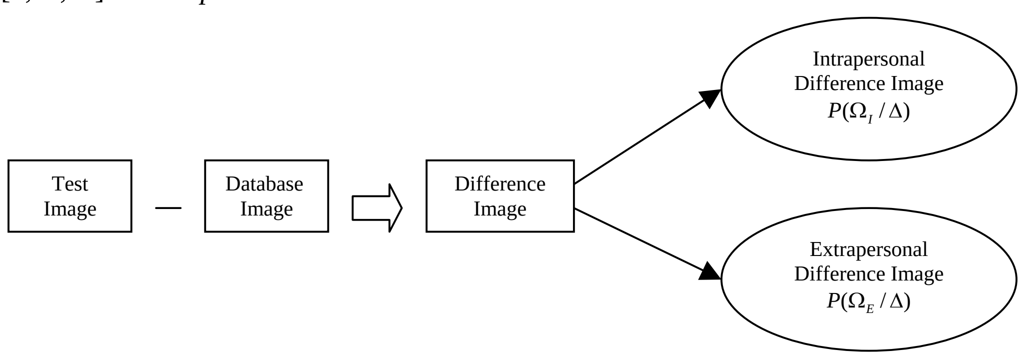


**Figure 1:** Classification of difference image between the test image and the database image into either of the 2 classes – Intrapersonal or Extrapersonal variations

We use a Bayesian i.e. maximum a posteriori system for the classification. Let $\Delta = I_1 - I_2$ denote the given intensity difference vector between two images having image intensity vectors $I_1$ and $I_2$. Let $\Omega_I$ and $\Omega_E$ denote the classes of intrapersonal and extrapersonal variations respectively. Then the posterior probability that $\Delta$ lies in the class of intrapersonal variations can be written according to Bayes rule as: -

$$P(\Omega_I / \Delta) = \frac{P(\Delta / \Omega_I) P(\Omega_I)}{P(\Delta / \Omega_I) P(\Omega_I) + P(\Delta / \Omega_E) P(\Omega_E)} \quad \text{………(1)}$$

Thus if $P(\Omega_I / \Delta) > P(\Omega_E / \Delta)$, or in other words $P(\Omega_I / \Delta) > \frac{1}{2}$ (since there are only 2 classes), we can say that there is a higher probability that the difference image lies in the class of intrapersonal variations i.e. the two images are of the same individual.

To calculate equation (1), we need to calculate estimates for both the intrapersonal and extrapersonal *likelihoods*, $P(\Delta / \Omega_I)$ and $P(\Delta / \Omega_E)$ respectively, as well as the two *priors* $P(\Omega_I)$ and $P(\Omega_E)$.

In their analysis [4], Moghaddam and Pentland assume that the two likelihoods are Gaussian distributed. Thus we can write the likelihood function as: -

$$P(\Delta / \Omega) = \frac{\exp\left(-\frac{1}{2} \tilde{\Delta}^T \Sigma^{-1} \tilde{\Delta}\right)}{(2\pi)^{N/2} |\Sigma|^{1/2}} \qquad (2)$$

where $N$ is the dimensionality of the difference vector i.e. the number of pixels in the image and $\Sigma$ is their covariance. $\tilde{\Delta} = \Delta - \overline{\Delta}$ is the mean subtracted difference image.

Following the derivation shown in [8], we can arrive at the following equation from equation (2)

$$P(\Delta / \Omega) = \frac{\exp\left(-\frac{1}{2} \sum_{i=1}^{N} \frac{y_i^2}{\lambda_i}\right)}{(2\pi)^{N/2} \prod_{i=1}^{N} \lambda_i^{1/2}} \qquad (3)$$

where the $y_i$ are the *principal components* and $\lambda_i$ are the *eigenvalues* obtained by running Principal Component Analysis (PCA) on the training data.

**Steps carried to get the values of $y_i$ and $\lambda_i$**

The training data consists of two matrices - the intrapersonal difference matrix, obtained by calculating difference images between images of the same individual from our training database, and the extrapersonal difference matrix, obtained by calculating the difference images between the images of different individuals.

During training, we run PCA twice, once for the set of intrapersonal difference images and once for the set of extrapersonal difference images. For each of the two sets, we obtain the following parameters: -

- A projection matrix - $\Phi$
- A vector of eigenvalues - $\lambda_i$

The values for $y_i$ are obtained by applying the projection matrix $\Phi$, to the mean subtracted difference image, i.e.

$y = \Phi^T \cdot \Delta$

**Basis Vector truncation:**

However in most cases, the dimensionality of the data *N*, is high. To build robust systems, we also need that the number of training images, which establishes an upper bound on the intrinsic dimensionality of the difference image subspace, be quite high. This makes the evaluation of equation (3), computationally expensive. As a result we truncate the PCA subspace, so as to retain only the *M* most significant dimensions. Here *M* is known as the *cutoff* parameter of the system.

Thus during training, we obtain a truncated projection matrix $\Phi_M$, that contains only the first *M* columns of $\Phi$ corresponding to the largest eigenvalues.

We can rewrite equation (3) as: -

$$P(\Delta / \Omega) = \frac{\exp\left(-\frac{1}{2} \sum_{i=1}^{M} \frac{y_i^2}{\lambda_i}\right)}{(2\pi)^{M/2} \prod_{i=1}^{M} \lambda_i^{1/2}} \cdot \frac{\exp\left(-\frac{1}{2} \sum_{i=M+1}^{N} \frac{y_i^2}{\lambda_i}\right)}{(2\pi)^{(N-M)/2} \prod_{i=M+1}^{N} \lambda_i^{1/2}}$$

Moghaddam and Pentland assume that the values of $\lambda_i$ are constant within the range

$M < i \leq N$. We calculate the average of the values of $\lambda_i$ in the range $M < i \leq N$ (which are available from PCA) and designate this $\rho$.

$$\rho = \frac{1}{N-M}\sum_{i=M+1}^{N}\lambda_i$$

Substituting $\rho$ for $\lambda_i$

$$P(\Delta/\Omega) = \frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{M}\frac{y_i^2}{\lambda_i}\right)}{(2\pi)^{M/2}\prod_{i=1}^{M}\lambda_i^{1/2}} \cdot \frac{\exp\left(-\frac{1}{2\rho}\sum_{i=M+1}^{N}y_i^2\right)}{(2\pi\rho)^{(N-M)/2}}$$

Now we define the *residual* $\varepsilon^2(\Delta)$ as

$$\varepsilon^2(\Delta) = \sum_{i=M+1}^{N}y_i^2 = \sum_{i=1}^{N}y_i^2 - \sum_{i=1}^{M}y_i^2 = \left\|\tilde{\Delta}\right\|^2 - \sum_{i=1}^{M}y_i^2$$

to obtain the final likelihood equation

$$P(\Delta/\Omega) = \frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{M}\frac{y_i^2}{\lambda_i}\right)}{(2\pi)^{M/2}\prod_{i=1}^{M}\lambda_i^{1/2}} \cdot \frac{\exp\left(-\frac{1}{2\rho}\varepsilon^2(\Delta)\right)}{(2\pi\rho)^{(N-M)/2}} \quad \ldots\ldots(4)$$

The choice of the two priors - $P(\Omega_I)$ and $P(\Omega_E)$ is an interesting one and will be addressed in the next section.

## 3 Experiments

For the above analysis, we need *frontal*, *geometrically aligned*, and *normalized* faces. For the purpose I used the Yale face database, which consisted of images, which were preprocessed accordingly. The training set consisted of 3 different images of the same individual – a normal image, an image without glasses & an image with center illumination – over a total of 14 individuals, making a total of 42 images.

The intrapersonal difference matrix was obtained considering the 3 difference images for each individual making a total of 42 difference images. The extrapersonal difference matrix was obtained by considering 70 random difference images between different individuals.

I considered 2 different estimates for matching –

- *Maximum Likelihood* estimate, considering only the evaluation of equation (4) for the class of intrapersonal variations
- *Maximum A Posteriori* estimate considering the full evaluation of equation (1)

### Maximum Likelihood analysis

The maximum likelihood (ML) analysis just evaluates the likelihood that the given difference image lies in the class of intrapersonal variations i.e. $P(\Delta/\Omega_I)$. That difference image which gives the maximum likelihood score can be said to be consisting of a match to the test image.

For training a ML classifier we only need to run PCA on the intrapersonal difference matrix. We thus obtain the various parameters needed to evaluate equation (4). From the likelihood scores we obtain, we need to set a certain threshold above which we can say that a match has been established. Picking up a single image corresponding to the maximum likelihood is also an option, but it will lead to an image being picked up even if there is no instance of the test image in our database. Unfortunately the threshold varies on a large scale depending upon the values of various parameters we set during training, size of training set etc. With $M$ = 12, and the training set of size 42 as mentioned above, I found a log-likelihood threshold of 350 as an optimal one i.e. those difference images with log-likelihood of more than 350 can be said to correspond to the same individual as the test image. Results averaged out for 50 runs of the ML classifier have been tabulated in Table 1.

### Maximum A Posteriori analysis

The Maximum a posteriori (MAP) analysis evaluates the value of the posterior probability that the difference image lies in the class of intrapersonal variations i.e. $P(\Omega_I/\Delta)$. Those difference images having $P(\Omega_I/\Delta) > \frac{1}{2}$ can be said to be corresponding to a match.

For training a MAP classifier we need to run PCA twice, once for the intrapersonal difference matrix and once for the extrapersonal difference matrix. We thus obtain 2 sets of parameters after training, which help us evaluate the likelihood scores. We also need to decide what priors to set for a correct analysis. Moghaddam and Pentland assume equal priors i.e. $P(\Omega_I) = P(\Omega_E)$ for their analysis. However during my testing I found that priors biased more towards the intrapersonal score perform significantly better than equal priors. Again depending on the training set size and the values of initial parameters, the optimal values of priors differ significantly. With a training set size of 42 intrapersonal & 70 extrapersonal difference images, and $M_I = 12$ & $M_E = 15$, I found priors in the ratio 3:1 towards $\Omega_I$ i.e. $P(\Omega_I) = 0.75$ and $P(\Omega_E) = 0.25$, work better. This fact though counterintuitive, may arise due to the fact that there is expected to be more bias for any difference image to have a higher extrapersonal likelihood than intrapersonal likelihood, and our priors help in toning down this bias. The results averaged out for 50 runs of the MAP classifier are tabulated in Table 1.

## 4 Results and Discussion

| | ML estimator | MAP estimator |
|---|---|---|
| Test image – Internal | 95.45% | 94.33% |
| Test image – External | 84.89% | 86.33% |

**Table 1: Classification accuracy**

The Yale database has images of 15 individuals out of which 14 were used for training. For testing purposes I ran the 2 classifiers considering 2 different conditions:

1. When the test image was randomly chosen from the 42 images, which were used to train the classifier,

2. When the test image was randomly chosen from the views of the 14 individuals but under different conditions (lighting, pose etc.) than the training images, or among the different views of the 15$^{th}$ person.

The results show us that both the classifiers give us reasonable performance, with high degree of accuracy. It seems that when the test image is internal, the ML classifier actually outperforms the MAP classifier. This is in accordance to the results of Moghaddam and Pentland who found that the ML classifier under certain conditions gives better results than the MAP classifier. This coupled with the fact that the ML classifier takes about half the time of the MAP classifier may hint at the fact that the ML classifier is a better choice when speed is our main concern or we are dealing with very large databases.

This method is a simple, elegant way to solve the problem of face recognition. The only problems with this method, which I found were that some of the global variables (like the priors, or the values of *M*'s) do not have any closed-form expression from which they can be calculated. Deciding on optimal values of these parameters at this point of time seems to be highly dependent on manual intervention.

However this method has a number of significant advantages. It is easily scalable to large databases as we consider only a certain (*M*) number of principal components, which reduces the training time to a large extent without affecting the accuracy too much. To build bigger and more robust systems we will have to have a good training set of images which incorporates different conditions in lighting, pose etc. Also as the results of this experiments show, it will be well advisable to use a maximum likelihood analysis when it would suffice to do so.

**References:**

[1] B. Moghaddam, T. Jebara, A. Pentland, "Bayesian Face Recognition", *Pattern Recognition*, Vol 33, Issue 11, pps 1771-1782, November 2000.

[2] W. Y. Zhao, R. Chellappa, A. Rosenfeld, P. J. Phillips, "Face recognition: A literature survey", *UMD CAR Technical Report* CAR-TR-948, 2000.

[3] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation", *IEEE transactions on Pattern Analysis and Machine Intelligence*, PAMI-19(7):696-710, July 1997.

[4] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object detection", In *IEEE Proceedings of the Fifth International Conference on Computer Vision (ICCV'95):*786-793, June 1995.

[6] R. Brunelli, and T. Poggio(1993*)*,“Face Recognition: Features versus Templates”, *IEEE Transactions on PAMI*, 15(10):1042-1052.

[7] M. Turk and A. Pentland, "Eigenfaces for recognition" *J. Cognitive Neuroscience*, vol. 3, pp. 71-86, 1994.

[8] L. T. Teixeira, “The Bayesian Intrapersonal/Extrapersonal Classifier”, *Master’s Thesis*, Colorado State University, 2003.

[9] B. Moghaddam , W. Wahid , A. Pentland, “Beyond Eigenfaces: Probabilistic Matching for Face Recognition”, *Proceedings of the 3rd. International Conference on Face & Gesture Recognition*, p.30, April 14-16, 1998.