Gaussian Mixture Models based on the Phase Spectra for Illumination Invariant Face Identification on the Yale Database

Sinjini Mitra and Marios Savvides

Abstract—The appearance of a face is severely altered by illumination conditions that makes automatic face recognition a challenging task. In [14], we introduced an illumination-invariant face identification method based on Gaussian Mixture Models (GMM) and the phase spectra of the Fourier Transform of images. In this paper we explore the application of this identification scheme on the Yale database that contains images with a greater degree of illumination variations. The novelty of our approach is that the model is able to capture the illumination variations so aptly that it yields satisfactory results without an illumination normalization unlike most existing methods. Identification based on a MAP estimate achieves misclassification error rate of 3.5% and a low verification rate of 0.4% on this database with 10 people and 64 different illumination conditions. Both these sets of results are significantly better than those obtained from traditional PCA and LDA classifiers. We next show that upon illumination normalization, our method succeeds in attaining near-perfect results using the reconstructed images. A rigorous comparison with existing state-of-the-art approaches demonstrates that our proposed technique outperforms all of those. Furthermore, some statistical analyses pertaining to Bayesian model selection and large-scale performance evaluation based on random effects model are included.

I. INTRODUCTION

The significance of face as a biometric is ever increasing today in order to ensure the security of many establishments, from the recording of biometric information (photo and fingerprint) of foreign passengers at all U.S. airports (the US-VISIT program) to surveillance cameras in bank ATMs. As opposed to feature-based approaches, model-based systems are becoming popular owing to their ability to accurately capture variability in the data. Some well-known model-based approaches include Gaussian models ([21]), deformable models ([25]), and the inhomogeneous Gibbs models ([10]) that are particularly good at capturing the local details of a face using a minimax entropy principle. One class of flexible statistical models is Mixture Models ([12]). These models can represent complex distributions through an appropriate choice of its components to represent accurately the local areas of support of the true distribution. Apart from statistical applications, Gaussian Mixture Models (GMM), the most popular of the mixture models, have also been used in computer vision. [26] used GMM for modeling the shape and texture of face images.

S. Mitra is with the Information Sciences Institute, University of Southern California, Marina del Rey, CA 90292 mitra@isi.edu
M. Savvides is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213. msavvid@cs.cmu.edu

The problem of human face identification under illumination variations is also well-researched ([4], [9]). [5] proposed an appearance-based algorithm for face recognition across pose by estimating the eigen light-field from a collection of images. [8] developed a bilinear model of an illumination subspace given arbitrary shape parameters from a 3D face model. Although all these approaches to devising face recognition systems are based on the spatial domain pixel intensities, recently much research effort has focused on the frequency domain as well, whose useful properties have been successfully exploited in various signal processing applications ([15]). The frequency spectrum of an image consists of two components, magnitude and phase. In 2D images particularly, the phase captures more of the image intelligibility than magnitude and hence is very significant for performing image reconstruction ([6]). [18] showed that correlation filters built in the frequency domain can be used for efficient face identification, and recently, [16] proposed correlation filters based only on the Fourier domain phase which perform as well as the original filters. [17] demonstrated that performing PCA in the frequency domain using only the phase spectrum outperforms spatial domain PCA and is robust to illumination variations.

In [13] we first introduced a novel GMM-based face identification system in the frequency domain by exploiting the significance of the phase spectrum, and applied it to the CMU-PIE database ([20]). Unlike many face recognition techniques ([2], [11], [1], [24]), our method does not require illumination normalization, nor does it require identifying the direction of the light source in advance, and can be trained with any subset of the original images without requiring a representation of all possible illumination states. This paper explores this method in more depth by applying it to the Yale database that contains images under more diverse illumination conditions. Upon performing illumination normalization on the images using the Quotient Image Relighting ([19]) approach, our approach yields highly accurate results with the re-constructed images that are superior to many existing state-of-the-art methods as seen from a rigorous comparative study. Finally, we apply a random effects model framework to evaluate the performance of this method for large-scale databases ([14]). This will help assess how well our method works in practical applications such as, in the airports where hundreds of thousands of people pass every day. Our evaluation protocol will help answer the following questions:

- How do different image properties affect the match score of a biometric system in the general popula-
tion, and what are the predicted score distributions for authentics and impostors, based on these image properties?

- What are the predicted authentic and impostor match score distributions and what error rates (both FAR and FRR) can be expected when a certain biometric system is applied to a large unknown database?

The rest of the paper is organized as follows. Section II presents a brief description of the GMM approach for human identification along with some results. Section III discusses the comparison with current approaches after performing illumination normalization on the images. We then present the large-scale performance evaluation method and results in Section IV, and conclude with a discussion in Section V.

II. GMM-BASED IDENTIFICATION METHOD

Because of their ability to capture heterogeneity in a cluster analysis context, finite mixture models provide a flexible approach to the statistical modeling of a wide variety of random phenomena. As any continuous distribution can be approximated arbitrarily well by a finite mixture of normal densities, Gaussian mixture models (GMM) provide a suitable semiparametric framework for modeling unknown and complex distributional shapes. Mixtures can thus handle situations where a single parametric family fails to provide a suitable semiparametric framework for modeling unknown and complex distributional shapes. Mixtures can thus handle situations where a single parametric family fails to provide a satisfactory model for local variations in the observed data, and offer the scope of inference at the same time.

A. The Model Framework

Let \( (Y_1, \ldots, Y_n) \) be a random sample of size \( n \) where \( Y_j \) is a \( p \)-dimensional random vector with probability distribution \( f(y_j) \) on \( \mathbb{R}^p \). Also, let \( \theta \) denote the vector of the model parameters. We write a \( g \)-component mixture model in parametric form as:

\[
f(y_j; \Psi) = \sum_{i=1}^{g} \pi_i f_i(y_j; \theta_i),
\]

where \( \Psi = (\pi_1, \ldots, \pi_g, \theta_1, \ldots, \theta_g)^T \) contains the unknown parameters, \( \theta_i \) represents the model parameters for the \( i^{th} \) mixture component and \( \pi = (\pi_1, \ldots, \pi_g)^T \) is the vector of the mixing proportions with \( \sum_{i=1}^{g} \pi_i = 1 \). When the mixture components have a multivariate Gaussian distribution, each component is given by:

\[
f(y_j; \theta_i) = \frac{1}{2\pi|\Sigma_i|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} (y_j - \mu_i)^T \Sigma_i^{-1} (y_j - \mu_i) \right\}
\]

where \( \phi \) denotes the multivariate Gaussian density with mean vector \( \mu_i \) and covariance matrix \( \Sigma_i \), \( i = 1, \ldots, g \). The mixture model now has the form:

\[
f(y_j; \Psi) = \sum_{i=1}^{g} \pi_i \phi(y_j; \mu_i, \Sigma_i).
\]

Despite the significance of phase in face identification tasks ([6]), modeling the phase angle poses several difficulties such as, the "wrapping around" property (it lies between \(-\pi \) and \( \pi \)) and its sensitivity to distortions (such as illuminations). This prompted us to use the real and imaginary parts of the frequencies in the "phase-only" spectrum as an alternative representation of the phase. This is a simple yet effective way of modeling phase since it does not suffer from the difficulties associated with direct phase modeling.

Let \( R_{s,t} \) and \( I_{s,t} \) respectively denote the real and the imaginary part at the \((s, t)^{th}\) frequency of the phase spectrum of the \( j^{th} \) image from the \( k^{th} \) person, \( s, t = 1, 2, \ldots, 100 \), \( k = 1, \ldots, 65 \), \( j = 1, \ldots, 21 \). We model each frequency separately and for each individual; that is, we model \( Y_{s,t} = (R_{s,t}, I_{s,t}) \), \( j = 1, \ldots, n \) where \( n \) denotes the total number of training images, as a mixture of bivariate normal distributions, the mixture components being given by

\[
\phi(y_{s,t}^{k}; \mu_{s,t}^k, \Sigma_{s,t}^k) = N[\mu_i, \Sigma_i], \quad i = 1, \ldots, g
\]

where \( \mu_i \) and \( \Sigma_i \) are respectively the frequency-wise mean vector and variance-covariance matrix, whose elements form the unknown parameters of \( \Psi \). The mixture model for each frequency for each person can then be written as:

\[
f_{s,t}^{k}(y_{s,t}; \Psi) = \sum_{i=1}^{g} \pi_{i,s,t}^{k} \phi(y_{s,t}^{k}; \mu_{i,s,t}^k, \Sigma_{i,s,t}^k), \quad j = 1, \ldots, n.
\]

It is a commonly known fact in signal processing that an image of good quality and identifiability can be reconstructed using only few low frequency components around the origin ([15]). So we model only the low frequencies within a \( 40 \times 40 \) square grid region (determined by experimentation) around the origin of the spectral plane. Moreover, owing to the Hermitian symmetry of the frequency components, only half of these frequencies need to used in the model (real part is symmetric and imaginary part is anti-symmetric). This succeeds in reducing dimensionality considerably. The complete model for each person is then given by (based on the independence assumption):

\[
f^{k}(y) = \prod_{s=1}^{S} \prod_{t=1}^{T} f_{s,t}^{k}(y_{s,t}; \Psi).
\]

The mixture components are expected to represent the different illumination levels in the images of each person. Recall that there are 64 different illumination conditions in the images of a person in the Yale database.

We use a Bayesian estimation technique using Gibbs Sampler for estimating the unknown model parameters. This yields a Markov chain \( \{\Psi[k], k = 1, 2, \ldots\} \) whose distribution converges to the true posterior distribution of the parameters. The point estimates of the parameters are formed by the posterior means, estimated by the average of the first \( N \) values of the Markov chain. To reduce error associated with the fact that the chain takes time to converge to the correct distribution, however, we discard the first \( N_1 \) samples as burn-in. Thus our estimates are given by

\[
E(\Psi | y) = \sum_{k=N_1+1}^{N} \Psi[k] \frac{N - N_1}{N_1}.
\]

Details of the estimation process can be found in [13].
B. Classification and Verification

We use a MAP (maximum a posteriori) estimate based on the posterior likelihood to classify the test images. For a new observation $X = (R_{s,t}, I_{s,t}, s, t = 1, \ldots, 100)$ extracted from the phase spectrum of a test image, we can compute the likelihood under the model for person $k$ by evaluating

$$h(X|k) = f^k(x),$$  \hspace{1cm} (8)

where $f^k(\cdot)$ is as in Equation 6. The posterior likelihood of the test image belonging to person $k$ is then given by:

$$f(k|X) \propto h(X|k)p(k),$$ \hspace{1cm} (9)

where $p(k)$ denotes the prior probability for each person which we assume to be uniform. A particular image will then be assigned to class $C$ if:

$$C = \arg \max_k f(k|X).$$ \hspace{1cm} (10)

For computational convenience, it is a convention to work with log-likelihoods in order to avoid numerical overflows/underflows in the evaluation of Equation 8.

C. Results on The Yale Database

The “Yale Face Database B” ([4]) consists of images of 10 individuals under 64 different illumination conditions, some sample images being shown in Figure 1. These images are cropped and aligned, the final images being of dimension $90 \times 80$. Owing to the properties of our method, we do not need to treat the 4 subsets of the database separately as done by many researchers.

![Sample images of 2 people (along the columns) from the Yale-B database under six different illumination conditions (along the rows).](image)

We start with $g = 2$ and use a total of $N = 5000$ iterations of the Gibbs sampler allowing a burn-in of $N_1 = 2000$ for each frequency. We select different training and test sets for each person to study how the number of training images affect the classification results. The training images in each case are randomly selected and the rest used for testing. Furthermore, this selection of training set is repeated 20 times (in order to remove selection bias) and the final errors are obtained by averaging over the 20 iterations. Table I summarizes the classification results. The standard deviations of the error rates over the 20 repetitions show that results are fairly stable across training/test sets.

These results illustrate that our proposed method is effective in dealing with a large number of illumination conditions. This is particularly attractive given the fact that no illumination normalization is performed prior to classification, and both training and testing are done on the original (cropped) images with varying illumination.

D. Model Selection

Often we can improve model fit and classification accuracy by using a greater number of mixture components. Recall that, in our models, the mixtures represent the different illumination conditions in the images of a person. It is therefore reasonable to expect that a larger training set is likely to have more variations with respect to the nature of illumination, and this may be better captured with more mixture components in the form of efficient clustering.

The results with $g = 3$ and $g = 4$ are included in Table II. We observe significant improvements in the classification performance for both the cases with $g = 3$ (p-value $< 0.001$). However, significant improvement is limited when the number of mixture components was further increased to $g = 4$ (p-value $> 0.4$). Therefore, $g = 3$ can be taken to be optimal for this dataset. We note here that a GMM with $g = 1$ (the traditional bivariate Gaussian model with no
mixture components) performed really poorly in all cases, yielding misclassification rates of over 50%. This shows that a single Gaussian distribution is unable to capture the inherent variability present in the different images due to illumination effects.

### Table II

**Error rates with more mixture components.**

<table>
<thead>
<tr>
<th>Case no.</th>
<th>Error Rate(g=3)</th>
<th>Std. dev</th>
<th>Error Rate(g=4)</th>
<th>Std. dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case (i)</td>
<td>2.34%</td>
<td>0.56%</td>
<td>2.10%</td>
<td>0.43%</td>
</tr>
<tr>
<td>Case (ii)</td>
<td>3.87%</td>
<td>1.02%</td>
<td>3.15%</td>
<td>0.98%</td>
</tr>
</tbody>
</table>

As with any statistical modeling experiment, it is desirable to use a pre-specified criterion for model selection so that no heuristics are involved in the optimal choice. We use Bayes Factors as our model selection criterion for the mixture models, a popular Bayesian model selection technique particularly for MCMC-based methods ([7]). For data D assumed to have arisen under one of two hypotheses $H_1$ and $H_2$ according to a probability density $p(D|H_1)$ or $p(D|H_2)$, the Bayes Factor is defined as:

$$B_{12} = \frac{p(D|H_1)}{p(D|H_2)}.$$  

(11)

Methods to calculate Bayes factor include asymptotic approximation, posterior simulation, MCMC, importance sampling, and so on (details can be found in [7]).

Model selection in our case involves choosing the number of mixture components $g$. The three hypotheses of interest are $H_1: g = 2$ and $H_2 : g = 3$, $H_3 : g = 4$. The Bayes Factors for comparing these hypotheses are shown in Table III ($2\log_{10} B_{ij}, i, j = 1, 2, 3$). Following the interpretation of Bayes Factors outlined in [7], we conclude that the data show quite strong evidence that both $H_2$ and $H_3$ should be preferred over $H_1$. However, the evidence in support of $H_3$ over $H_2$ is not that significant. Moreover, using more components introduces many additional parameters which increases the consumption of computing resources. Thus, a GMM with $g = 3$ best fits the training data and has the best classification performance in terms of optimizing the error rate and computation complexity, it being consistent with the classification results shown above.

### Table III

**Bayes factors for the 3 model comparisons.**

<table>
<thead>
<tr>
<th>$H_1$ vs. $H_2$</th>
<th>$H_1$ vs. $H_3$</th>
<th>$H_2$ vs. $H_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.23</td>
<td>4.89</td>
<td>1.55</td>
</tr>
</tbody>
</table>

E. Performance Comparison

This section presents a brief comparative study of the identification performance of our GMM-based method with those from some existing methods applied to the same subset of the Yale database. The competing classifiers used are Principal Components Analysis (PCA) and Linear Discriminant Analysis (LDA), the results being included in Table IV. We use $g = 3$ and 20 training images. As we can see clearly, both the traditional PCA and LDA classifiers fail completely for both datasets, yielding high misclassification error rates, and our proposed GMM-based method outperforms them by an overwhelming margin.

### Table IV

**Comparing classification and verification error rates of our proposed method with traditional classifiers.**

<table>
<thead>
<tr>
<th></th>
<th>GMM (our method)</th>
<th>PCA</th>
<th>LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification</td>
<td>3.34%</td>
<td>33.43%</td>
<td>43.21%</td>
</tr>
<tr>
<td>Verification</td>
<td>0.20%</td>
<td>29.01%</td>
<td>26.72%</td>
</tr>
</tbody>
</table>

III. APPLICATION TO “NORMALIZED” IMAGES

Recall that the Yale database contains 64 different illumination conditions representing the different directions of the available light sources. Now for a test image, once its illumination level is determined by the classification process, that information can be used to remove or reduce the effect of illumination, a process commonly referred to as “illumination normalization”. We use a top-of-the-shelf normalization technique, called the Quotient Illumination Relighting (QIR) method ([19]), which is used to synthesize images under a pre-defined uniform lighting condition (the canonical form) from the provided face images captured under non-uniform lighting condition. It yields a good representation if an image with uniform illumination is available. Recognition can then be performed using these reconstructed images. We do not include much details about the procedure in this paper, and an interested reader is referred to [19].

A. Performance comparison using reconstructed images

We now compare the performance of our proposed technique when applied to the normalized or reconstructed images, with other existing state-of-the-art methods that are known to perform satisfactorily on illumination-free images ([12], [11], [1], [22] and [24]). Since all these methods concentrate on human classification, we compare only classification performance (and not verification performance). Table V shows these results from the various methods. The LDA and PCA classifiers had failed completely when using the original images with illumination and produced error rates close to random guessing. With the illumination effect removed, both these well-known methods proved to be more successful with less than 4% error rate. Our GMM-based approach also yielded improved results when applied to the illumination-free images, and outperforms all the other methods. Another interesting thing to note here that, [11] obtained an error rate of 25.6% when using the raw images of the Yale database (that is, with illumination variation present in them), which is far worse than our result of 2.34% error rate on the same database using the same set of raw images. As a side note, although we did not compare the verification performance, our GMM-based approach produced perfect results (0% error rate) on the normalized images from the two datasets. These results, therefore, establish the efficacy of our GMM-based classifier in the frequency domain, thus leading to very good
IV. LARGE-SCALE PERFORMANCE EVALUATION

Traditional performance evaluation tools like linear regression models, ANOVA ([23]) are fixed effects models and the inference from them cannot be generalized beyond the people in the database at hand. This difficulty is obviated by the use of random effects models ([3]) which provide a flexible framework for extending inference based on the current database to a bigger population by assuming that the particular subset of subjects in the current database is a random sample from a bigger population. The regression framework allows the inclusion of any number of potential covariates representing image properties (distortions, clarity) and soft biometric traits (age, gender), and so on, that may influence the system performance, and quantitatively determine the extent to which these factors affect performance in a general population. We introduced this performance evaluation framework in [14] and apply this now to the results from the GMM-based classifier on the Yale database.

A. The Model Framework

For the sake of completeness, we provide a brief description of the model here. Let $Y_{ij}$ denote the outcome for the $j^{th}$ observation on the $i^{th}$ subject in the database, while $x_{ij}^{(m)}$ denotes the corresponding value for covariate $m$. We adopt the following hierarchical model:

$$Y_{ij} \sim_{ind.} N(\theta_i + \sum_{m=1}^{M} \beta_i^m x_{ij}^{(m)}, \sigma^2), \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i,$$

(12)

where $M$ is the total number of covariates in the study. We assume that the slope-intercept vectors for each individual are drawn from a common population, thus taking into account the heterogeneity across individuals:

$$\theta_i \equiv (\alpha_i, \beta_i^1, \ldots, \beta_i^M)^T \sim MVN(\theta_0 \equiv (\alpha_0, \beta_0^1, \ldots, \beta_0^M)^T, \Sigma),$$

(13)

$i = 1, \ldots, k$. We then select conjugate hyper-priors for the other parameters and use the Gibbs Sampler to estimate the unknown parameters by simulating from the respective posterior distributions. Inference from this model is based on the marginal posteriors for the population parameters $\theta_0 = (\alpha_0, \beta_0^1, \ldots, \beta_0^M)$ and posterior predictive distributions $p(y_j | y)$. This is obtained with Gaussian kernel density estimates for new data generated using Equation 12 with the post-convergence values of the parameters. Details about the model and inference procedures can be found in [14].

B. Application

We apply the random effects methodology to the verification results from the Yale database based on the un-normalized images since we obtained perfect verification with the normalized images. The response variable in this case is posterior log-likelihood, the authentication score. We use two covariates, one denoting authenticity and another representing the illumination level (recall, there are 64 different illumination conditions in the Yale database). Table VI shows the estimates of the population parameter $\theta_0$ which quantitatively show the effect of the covariates on the scores.

We see that authenticity has a significant effect on authentication score while illumination has no effect on match scores. This is also evident from the predictive intervals (confidence intervals based on the posterior distributions) — the one for illumination contains 0 while that for authenticity does not. This shows that our GMM-based classifier is robust to the illumination changes in the test images.

The posterior predictive distribution of the authentication scores are shown in Figure 3. We generated 1000 values for each system and 500 each for the authentications and the impostors. As can be seen, there exists a clear separation among the predicted values of authentic and impostor people; in fact, the distributions of the score statistics appear to be a mixture of two distinct distributions. The amount of overlap in the tails of the authentic and impostor distributions indicates the chances of false alarm and this clearly shows that the GMM method has a considerably low risk of errors (both false positive and false negative).

![Fig. 3. Predictive posterior distribution of authentication scores from the GMM-based classifier. X-axis: posterior log-likelihood.](image-url)
V. DISCUSSION

In this paper, we have established the efficiency of our proposed GMM-based classification method based on the the phase spectra. Modeling phase with the help of an appropriate representation of its variability across different images of a person is indeed a challenging task and our experiments show that our proposed model is able to handle it fairly well. We point out again that the novelty of our method lies in the fact that the human identification method does not need to rely on removing the lighting effects in advance like many existing work, and works directly on the raw images with extreme illumination variations. Not only this, we have demonstrated that our approach has a fairly general framework and we are exploring applications of our methodology to images with expression and pose variations. Thus this should be useful in practice for handling real life databases that are often subject to extraneous variations. Very good identification and verification error rates are obtained both on images with and without illumination effects that are superior to those from current state-of-the-art methods, as well as to traditional classifiers like LDA and PCA. In conclusion, both GMM and phase have enormous potential in computer vision, and harnessing this combined strength has indeed proved to be a success. The large-scale evaluation framework also help understand the actual potential of the method when applied in real-life situations.

Our immediate future direction of work consists of assessing the performance of our method on more diverse databases with expression and pose variations and also on other biometric traits, such as iris and fingerprints.