Unconstrained Face Recognition using MRF Priors and Manifold Traversing

Ricardo N. Rodrigues, Greyce N. Schroeder, Jason J. Corso and Venu Govindaraju

Abstract— In this paper, we explore new methods to improve the modeling of facial images under different types of variations like pose, ambient illumination and facial expression. We investigate the intuitive assumption that the parameters for the distribution of facial images change smoothly with respect to variations in the face pose angle. A Markov Random Field is defined to model a smooth prior over the parameter space and the maximum a posteriori solution is computed. We also propose extensions to the view-based face recognition method by learning how to traverse between different subspaces so we can synthesize facial images with different characteristics for the same person. This allow us to enroll a new user with a single 2D image.

I. INTRODUCTION

One of the remaining main challenges in face recognition research nowadays is to develop robust recognition in unconstrained environments. In such environments, characteristics such as pose, ambient illumination, facial expression, etc. cannot be controlled[1] (see figure 1). However, it is known that traditional face recognition methods like Eigenfaces [10], Fisherfaces [1] and Elastic Bunch Graph Matching [12], are not robust against high variations in the image characteristics. Several techniques have been proposed to make face recognition more robust against variations. These techniques can be divided in three groups:

Fig. 1. Example of facial images with different characteristics: frontal, 40° rotation, smiling and illumination variation.

- **3D based models:** These methods try to extract a 3D model for the face, which is invariant to pose variations and contain more information about the subject identity than a 2D image. A 3D face model can be extracted using a 3D scanner or using morphable models [4], which extract a 3D face model based on a 2D face image. The drawback with these methods is that they tend to be more expensive when compared with 2D

methods (either due to the need of specific 3D hardware or/and due to a higher computational burden).

- **Non-linear based models:** Variations in the face pose, illumination and expression tend to cause highly non-linear variations in the respective 2D face image. The face recognition methods in this group use non-linear techniques to model facial variations. For example, [13] apply kernel principal component analysis and [14] use locally linear embedding to learn a non-linear image subspace.

- **Pre-processing:** Given an image with any characteristics, these methods try to apply image transformations techniques with the objective of generating an equivalent image in normal conditions (e.g. frontal view, neutral expression, neutral lighting, etc). The matching between two images with different characteristics is done by first transforming both images into the same characteristic and then computing a distance between them. For example, [5] propose a method that apply a image preprocessing algorithm that compensates for illumination variations in images. Beymer *et al* [3] apply a method to rotate a frontal image. In [15], the authors present a method that use Light-Fields and Lambertian reflectance to model the distortions provoked by different poses and illumination. These methods usually treat each type of variation independently, and the development of a general method that can deal with several types of variations at the same time have not been proposed yet.

In this paper, we investigate new methods to improve the modeling of facial variations. We adopt a multiple subspace model similar to the view-based method of [9], where each facial pose is modeled by a different eigenspace. This model relates with the non-linear based models described above since the several "local" eigenspaces (one for each facial pose) implicitly form a global non-linear manifold of facial images. We propose two main improvements to this method:

- Traditionally, the parameters (mean and covariance) for the subspaces are estimated independently for each characteristic. However, the independence assumption may be over simplistic for some types of characteristic variations, leading to poor parameter estimation, specially when few samples are available as training data. We present a method where the assumption of independence between pose variations is relaxed. The underlying assumption for our model is that similar facial poses tend to have similar parameters. Hence,

[1]In this paper, we refer to the pose, illumination and expression of a face image as the image characteristic

we define a Markov Random Field (MRF) to impose a smoothness constrain over the parameters of similar poses. We show that this smoothness constrain improves the estimation for the parameters.

- In the original method [9], several images with different characteristics are needed for each enrolled person. However, in practice, it is desired to enroll a person with one single 2D image. We present a method that overcomes this problem by learning how to traverse between different subspaces. Thus, given a single image from a person, we can generate simulated images from the same person but with different characteristics. This transformation from one characteristic to another allows us to match two images as they had the same characteristic, similarly to the pre-processing methods described above. The advantage of our method is that the same basic model can be applied to model any kind of facial variation (e.g. pose, expression, illumination, aging, occlusion, etc).

In section II, we briefly describe the view-based method of [9] and introduce the notation used in this paper. In section III, we present the MRF used to improve the parameters estimation of subspaces with pose variation. Some experiments on this method are also presented in this section. Section IV describes in detail how to traverse between two subspaces with different characteristics. Finally, in Section V we discuss the conclusions of our work and give directions for future work.

II. PRELIMINARY CONCEPTS: THE CHARACTERISTIC-BASED EIGENSPACE REPRESENTATION

Let $\mathbf{I}$ represent a vectorized facial image and $q \in Q$ represent the characteristic of an image (e.g. pose angle, illumination, expression). Note that the space of possible facial characteristics Q can be represented by a orthogonal space where each axis correspond to one characteristic. Let $\mathbf{I}_q$ be an image with characteristic q and suppose we have a set of N training images $\mathcal{D}_q = \{\mathbf{I}_q^{(i)}\}_{i=1}^{N}$ belonging to different persons but with the exactly same characteristic q. The Principal Component Analysis (PCA) method can be applied on the training set of images to compute a subspace $\mathbf{P}_q$ with dimension M ($M \ll |\mathbf{I}_q|$) that captures the principal modes of variation as follows:

$$\mu_q = \frac{1}{N} \sum_{i=1}^{N} \mathbf{I}_q^{(i)} \tag{1}$$

$$\Sigma_q = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{I}_q^{(i)} - \mu_q)(\mathbf{I}_q^{(i)} - \mu_q)^T \tag{2}$$

$$\mathbf{P}_q = \texttt{eigen}(\Sigma_q, M) \tag{3}$$

where μ_q is the mean of all images, Σ a covariance matrix and $\texttt{eigen}(\Sigma, M)$ is a function that returns the M eigenvectors corresponding to the M biggest eigenvalues for matrix Σ.

Since all images used to compute $\mathbf{P}_q$ have the same characteristics but belong to different persons, this subspace spans only the variations caused by different identities. Note that we are assuming all variations between two images with the same characteristics are caused exclusively by the difference in the identity of the two images.

The projection of an image $\mathbf{I}_q$ into the subspace $\mathbf{P}_q$ is computed as follows:

$$\tilde{\mathbf{I}}_q = (\mathbf{I}_q - \mu_q)\mathbf{P}_q \tag{4}$$

where $\tilde{\mathbf{I}}_q$ is a vector of size M that represent the projected image.

The matching between two images with the same characteristics can be done by first projecting both images in the subspace and then computing the distance between the two projected images. More formally, let $\mathbf{I}_q^{(1)}$ and $\mathbf{I}_q^{(2)}$ be two different face images with the same characteristic q. The distance between these two images in the subspace $\mathbf{P}_q$ is $d = \|(\tilde{\mathbf{I}}_q^{(1)} - \tilde{\mathbf{I}}_q^{(2)})\|$, where $\|\cdot\|$ denotes the Euclidean norm. Since the subspace $\mathbf{P}_q$ spans only the variation caused by different identities, the distance d can be interpreted as the distance between the respective identity of two images. A threshold operator can be applied to d in order to decide if both images have the same identity or not.

A problem with this method, that was not addressed in [9], arises when two images with different characteristics need to be matched. The two images cannot be projected in the same subspace because each subspace is trained for one specific characteristic. On the other hand, if the images are projected in different subspaces, the distance between them would not be exclusively related to variations in the identity. This problem is addressed in section IV.

III. IMPROVING THE LEARNING OF POSE EIGENFACES PARAMETERS WITH A SMOOTH CONSTRAIN

A. Method overview

Note that, in equations 1 and 2 the parameters $\theta_q = \{\mu_q, \Sigma_q\}$ are learned independently for each characteristic q. As stated before, this may not be ideal and may lead to poor parameter estimation. We propose a method that relaxes the independence assumption for pose variations. Let $q \in Q \subset [-90^o, 90^o]$ represent the pose angle of a facial image. Our model assumes that the parameters θ_q vary smoothly with respect to q. For example, our method assumes that the mean and covariance matrix of images with $q = 15^o$ should be similar to the mean and covariance matrix of images with $q = 0^o$ and $q = 30^o$. We use a prior distribution defined by a Markov Random Field (MRF) [6] over the parameter space to impose this smoothness constrain.

Let $Q = \{q_1, ..., q_K\}$ represent a finite set with K different angles and $\Theta = (\theta_{q_1}, ..., \theta_{q_K})$ represent the respective parameters for these pose angles. Without loss of generality, we assume Q is sorted in ascending order such that q_{k-1} and q_{k+1} are the two neighbors of q_k. Using the smoothness property, we define the probability of Θ as follows:

$$p(\Theta) = \frac{1}{Z} \exp[- \sum_{k=2}^{K-1} (\alpha_k \theta_{k+1} - \theta_k + (1 - \alpha_k)\theta_{k-1})^2] \tag{5}$$

where Z is the normalization constant and α_k is given by the relation between pose q_k to its neighbors and is defined as follows:

$$\alpha_k = \frac{q_k - q_{k-1}}{q_{k+1} - q_{k-1}} \quad (6)$$

When the angles in Q are equally separated we have $\alpha_k = 0.5$, $\forall q_k \in Q$ and 5 becomes an homogeneous MRF. Note that the probability function in equation 5 is maximum when the gradient for the parameters is constant (i.e. Θ is contained in a straight line on the parameter space).

As in the original eigenface method, the probability for an image $\mathbf{I}_q$ is modeled by a Gaussian:

$$p(\mathbf{I}_q|q,\theta_q) = \frac{1}{(2\pi)^{d/2}|\Sigma_q|^{\frac{1}{2}}} \exp[-\frac{1}{2}(\mathbf{x}-\mu_q)\Sigma_q^{-1}(\mathbf{I}_q-\mu_q)^T]$$
$$(7)$$

where $\theta_q = \{\mu_q, \Sigma_q\}$. The set of parameters for all pose angles is represented by $\Theta = \{\{\mu_k, \Sigma_k\}\}_{k=1}^K$.

Now, given a training set $\mathcal{D} = \{\mathcal{D}_{q_1}, ..., \mathcal{D}_{q_K}\}$, our objective is to find the maximum a posteriori (MAP) solution for the parameters as follows:

$$\Theta_{MAP} = \arg\max_{\Theta} p(\Theta|\mathcal{D}) \quad (8)$$
$$= \arg\max_{\Theta}\{\log p(\mathcal{D}|\Theta) + \log p(\Theta)\} \quad (9)$$

There is no closed form solution for equation 8, so we use Iterated Conditional Modes (ICM) [2] to solve it as follows:

1: Initialize all θ_k using (class independent) Maximum Likelihood.
2: **for** number of iterations **do**
3:    **for all** classes $k = \{1, ..., K\}$ **do**
4:       Optimize parameter θ_k by solving $\frac{d}{d\theta_k}p(\mathcal{D}|\Theta)p(\Theta) = 0$.
5:    **end for**
6: **end for**

It is important to say that the ICM algorithm converges locally. However, since the ML initialization provides a good initial guess, we conjecture that the global solution is been found.

B. Experiments

We use the FERET database [8] to perform the tests on our method. The FERET database is a public domain database containing face images with different characteristics. For our work, we use a specific set, referred as *b series*, containing 2200 face images collected from 200 people. This set of face images is composed by different subsets as characterized in Table I. Each subset contains exactly one image per person (what implies that each subset contain 200 images). In this section we use only subsets with pose variation (i.e. subsets ba to bi). Three experiments were conducted:

Experiment I: Parameter error - In our first experiment, we evaluate the learning rate for the ICM algorithm. Let $\hat{\Theta}_{ML}$ be the ML estimation (equations 1 and 2) using all 200 persons in the dataset as training. Let Θ_{ML} and Θ_{MAP} be

Subset	Pose angle	Description
ba	0	Frontal view
bb	+60	Subject face to his left
bc	+40	
bd	+25	
be	+15	
bf	-15	Subject face to his right
bg	-25	
bh	-40	
bi	-60	
bj	0	Alternative expression to ba
bk	0	Different illumination to ba

TABLE I
FERET *b series* FACE DATABASE.

the ML and MAP parameter estimation, respectively, using only 30 persons (randomly chosen) for training. Figure 2 shows the Euclidean distance from Θ_{ML} to $\hat{\Theta}_{ML}$ and from Θ_{MAP} to $\hat{\Theta}_{ML}$ for each iteration of the ICM algorithm. It can be seen that, after each iteration, Θ_{MAP} gets closer to the $\hat{\Theta}_{ML}$. Since we have used significantly more training data for learning $\hat{\Theta}_{ML}$ we conclude that, using the same amount of training data, Θ_{MAP} is a better estimation than Θ_{ML}.
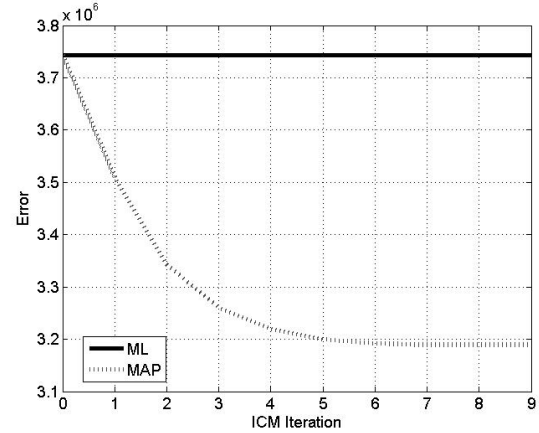

Fig. 2. MAP parameter error during ICM algorithm execution.

Experiment II: Reconstruction error - We use Θ_{ML} and Θ_{MAP} learned in the previous experiment to compare the image reconstruction error. Both sets of parameters are used to compute characteristic-based eigenspaces as described in section II. All images for the 170 persons not used in the training phase where projected into their respective eigenspaces and then reconstructed. The Euclidean distance (error) between the pixels of the original and projected image was computed for all images. The average error using the Θ_{ML} parameters was 259.2 and for the Θ_{MAP} parameters was 249.5. This shows that the subspaces computed using the MAP parameters are more accurate than the ML subspaces.

Experiment III: Pose classification - Now we consider the problem of facial pose estimation [7], where the objective is to classify the pose of a test facial image. Once again,

the parameters learned in the first experiment are used. Each image $\mathbf{I}$ in the test dataset is classified with the pose with the highest log-likelihood $\log p(\mathbf{I}|q, \theta_q)$. Note that the covariance matrix Σ_q has rank 29, so we assume unitary variation in the other dimensions. [2] An accuracy of 67.91% and 76.08% was obtained for the ML and MAP parameters, respectively.

Table II shows the gain obtained by the MAP parameters when compared to ML parameters.

Experiment	Improvement MAP
I	17.34%
II	3.89%
III	12.03%

TABLE II

IMPROVEMENT OF MAP PARAMETER ESTIMATION WHEN COMPARED TO ML PARAMETERS FOR ALL EXPERIMENTS.

Next, we describe our matching algorithm and evaluate the MAP parameters in a face recognition system.

IV. MATCHING BETWEEN IMAGES WITH DIFFERENT CHARACTERISTICS

A. Method overview

As stated before, the characteristic-based eigenfaces method does not provide a direct way of matching two images with different characteristics. To solve this problem, we propose a method where first each image is projected into the subspace of its respective characteristic. Then, using the same training data used to calculate the subspaces, we learn a mapping function that makes the mapping between the two subspaces. With this method, given an image with characteristic $q1$, we can obtain an estimation of this image with characteristic $q2$ such that the identity is preserved. When the transformation between two subspaces is linear, this method is related to the synthesis of facial images based on linear object classes [11].

Figure 3 illustrate the general idea. The space of possible characteristics is spanned by three axis: pose, expression and illumination. For each point in this space there is an image subspace (eigenspace) spanning the variations caused by different identities. Given training images from different persons for two different characteristics, we learn a transformation function that maps a point in one subspace to an equivalent point in the other subspace such that the identity is preserved.

B. Training

Let $\{\mathbf{I}_{q1}^{(i)}\}_{i=1}^N$ and $\{\mathbf{I}_{q2}^{(i)}\}_{i=1}^N$ represent two sets of training face images with characteristics $q1 \in Q$ and $q2 \in Q$, $q1 \neq q2$, such that image $\mathbf{I}_{q1}^{(i)}$ and $\mathbf{I}_{q2}^{(i)}$ belongs to the same person. The subspaces and means for each characteristic can be calculated independently using equations 1 and 3. Our

[2]This is equivalent to defining $\log p(\mathbf{I}|q, \theta_q) = recError_q(\mathbf{I}) + \log p(\mathbf{I}|q, P_q)$, where $recError_q(\mathbf{I})$ is the reconstruction error of image $\mathbf{I}$ when projected into subspace q and $p(\mathbf{I}|q, P_q)$ is a Gaussian in the subspace q with mean μ_q and diagonal covariance matrix given by the eigenvalues of Σ_q.
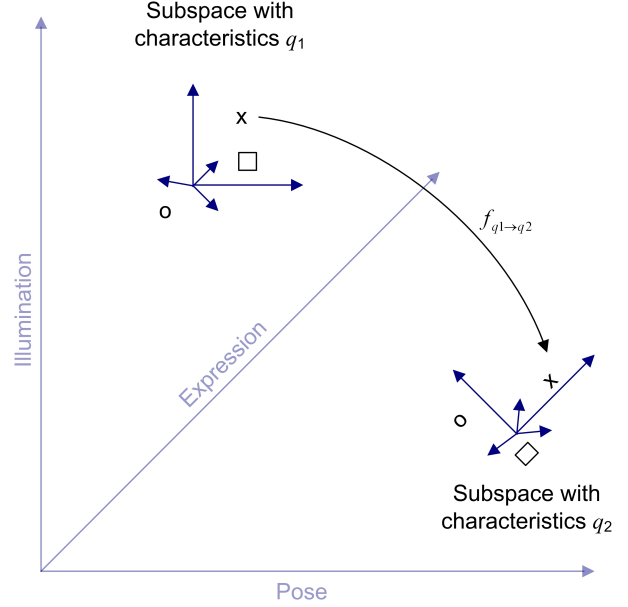

Fig. 3. Characteristic-based eigenspaces representation.

objective is to learn a transformation function that can map a projected face image with characteristic q_1 to a projected face image with characteristic q_2:

$$\tilde{\mathbf{I}}_{q2} = f_{q1 \to q2}(\tilde{\mathbf{I}}_{q1}) \tag{10}$$

The transformation function $f_{q1 \to q2}$ can be viewed as a regression function from one subspace to another. This function can be learnt by minimizing the following error:

$$\varepsilon = \sum_{i=1}^N \|\tilde{\mathbf{I}}_{q2}^{(i)} - \tilde{\mathbf{I}}_{q1} \mathbf{A}_{q1 \to q2}(\tilde{\mathbf{I}}_{q1}^{(i)})\| \tag{11}$$

In this paper we model the transformation function as a linear function, so the function in equation 10 can be represented as:

$$f_{q1 \to q2}(\tilde{\mathbf{I}}_{q1}) = \tilde{\mathbf{I}}_{q1} \mathbf{A}_{q1 \to q2} \tag{12}$$

where $\mathbf{A}_{q1 \to q2}$ is a square matrix of size M (subspaces dimensionality). Using equation 12, the minimum for equation 11 can be found by solving a linear system as follows:

$$\begin{pmatrix} \mathbf{I}_{q1}^{(1)} \\ \mathbf{I}_{q1}^{(2)} \\ \vdots \\ \mathbf{I}_{q1}^{(N)} \end{pmatrix} * \mathbf{A}_{q1 \to q2} = \begin{pmatrix} \mathbf{I}_{q2}^{(1)} \\ \mathbf{I}_{q2}^{(2)} \\ \vdots \\ \mathbf{I}_{q2}^{(N)} \end{pmatrix} \tag{13}$$

A point that needs to be stressed here is that images $\mathbf{I}_{q1}^{(i)}$ and $\mathbf{I}_{q2}^{(i)}$ are from the same person. This is essential to ensure that the transformation learned do not alter the face identity.

C. Transforming Images

Let $\mathbf{I}_{q1}$ be an face image from a person not included in the training set. We can transform this image from characteristic $q1$ to characteristic $q2$ following these three steps:

1) Project image $\mathbf{I}_{q1}$ into the subspace with characteristic $q1$;
2) Transform the projected image to subspace with characteristic $q2$;
3) Reconstruct the image using subspace with characteristic $q2$.

Figure 4 shows exactly how this process is done. Figure 5 show some examples obtained using the FERET-*series b* dataset. The first image for each user (frontal, neutral expression) is given and the other images are synthesized using our model. Visually, the results obtained are good. Note that the same basic model can be used to model all types of facial variations (e.g. expression, pose, illumination).
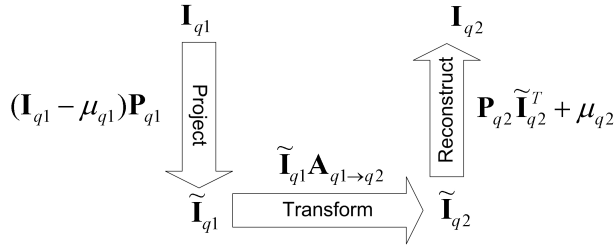
Fig. 4. Diagram showing the transformation from one image with characteristic $q1$ to another with characteristic $q2$.
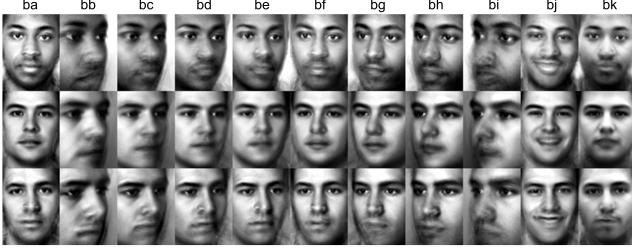
Fig. 5. Examples of image transformation. Given only the image in the first collumn, we generate an estimation for the images of the same person but with different characteristics. The generated images are shown in columns 2 to 10. Refer to Table I for the characteristic description of each image.

D. Matching

Let $\mathbf{I}_{q1}$ and $\mathbf{I}_{q2}$ represent two face images not in the training set. The matching between these two images can be done following these three steps:

1) Project each image into its respective subspace;
2) Transform one projected image to the other subspace;
3) Compute the distance between the projected images in the same subspace.

In the first and second steps, we use the classifier described in Experiment III of section III-B to find the characteristics

of a given image. Note that the distance in step 3 is computed using the projected images only, so there is no need to reconstruct any image. Moreover, this distance is computed within the same subspace, so we can say that it is related exclusively to the difference of identity between the images. The complexity of this matching algorithm is the same of the eigenfaces method since we only have one extra vector-matrix multiplication (step 2).

E. Experiments

In the following experiments we use 100 users (randomly chosen) as training data and the other 100 as testing. We compare the recognition rates for the ML and MAP parameters learned as described in section III. The standard eigenfaces method [10] is used as a baseline.

Experiment IV: Global face recognition - The *ba* image (frontal, neutral expression and neutral lighting) is used as the enroled image for each person in the test set. This image is matched against all other images in the test set and the Receiving Operating Curve (ROC) is computed. Despite the fact experiments I to III have shown the MAP parameters are better than the ML parameters, we did not see significant difference in the recognition rates (the improvements were on the order of 0.2%; for this reason, we report only the results using the MAP parameters). We conjecture this happened because the gains obtained by the improved parameters where not sufficient to significantly improve the discrimination between users. The results are shown in Figure 6. Nevertheless, it can be seen that our proposed matching algorithm present significantly better results when compared with the Eigenface method.
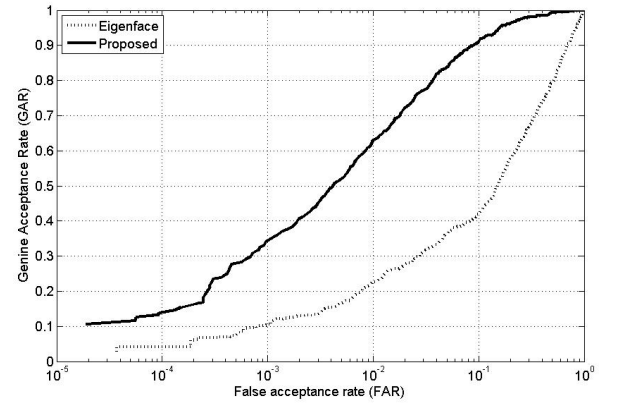
Fig. 6. Receiving Operating Curve (ROC) comparing the proposed method with MAP parameter estimation with the Eigenface method.

Experiment V: Subsets face recognition - In this experiment each characteristic is individually tested both for the verification task (where the objective is to discover if two images belong to the same person) as well as for the identification task (where the objective is to discover the identity of an image). The Equal Error Rate (EER) is reported

for the verification task and the identification Accuracy (ID-Acc) for the identification task. The results are shown in table III. We can note that the error rates are accentuated for bigger angle views.

	EER (%) Verification		ID-Acc (%) Identification	
Subset	Proposed	Eigenface	Proposed	Eigenface
ba	Enrolled image			
bb	13.9	42.6	50.0	6.0
bc	11.7	34.5	62.0	8.0
bd	8.7	27.3	86.0	20.0
be	6.4	20.7	94.0	48.0
bf	6.1	18.1	92.0	68.0
bg	8.1	28.0	78.0	20.0
bh	11.3	40.1	60.0	14.0
bi	18.2	44.9	36.0	6.0
bj	6.3	18.3	88.0	86.0
bk	8.7	27.9	82.0	50.0

TABLE III

EQUAL ERROR RATE (EER) (LOWER IS BETTER) AND IDENTIFICATION ACCURACY (ID-ACC) (HIGHER IS BETTER) FOR EACH SUBSET IN THE FERET DATASET.

V. CONCLUSION

The investigation presented in this paper has focused in two main topics: first, we have shown that we can obtain better estimation for the eigenspace parameters of facial images with different poses by using a MRF over the parameter space to model a smoothness assumption. Second, we extended the view-based method proposed in [9] so we are able to enroll an user with one single image and still perform the matching when a test image with different characteristics is presented.

By using the proposed MRF approach, we were able to reduce the image reconstruction error as well as improve the performance in a pose estimation application when compared with the traditional ML parameter estimation. However, no significant difference between the ML and MRF (MAP) parameter estimation was observed when these parameters were used in a face recognition application.

The characteristic-based face recognition presented in this paper has shown significantly better recognition rates when compared with the Eigenfaces method and has an equivalent computational complexity. However, in this paper, we have consider only test images with characteristics for which we have training data. Our future plans include methods that are capable of interpolating between subspaces as well as perform experiments with a wider range of facial variations.

REFERENCES

[1] Peter N. Belhumeur, Joao P. Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[2] Julian Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):259–302, 1986.

[3] D. Beymer and T. Poggio. Face recognition from one example view. *iccv*, 00:500, 1995.

[4] Volker Blanz. Face recognition based on a 3d morphable model. *Seventh IEEE International Conference on Automatic Face and Gesture Recognition*, 0:617–624, 2006.

[5] Ralph Gross and Vladimir Brajovic. An image preprocessing algorithm for illumination invariant face recognition. *Audio- and Video-Based Biometric Person Authentication*, 2688/2003:1055, 2003.

[6] Stan Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer, 3rd edition, 2009.

[7] E. Murphy-Chutorian and M.M. Trivedi. Head pose estimation in computer vision: A survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(4):607–626, April 2009.

[8] NIST. The facial recognition technology (feret) database. Technical Report NISTIR 7151, National Institute of Standards and Technology (NIST), August. http://www.itl.nist.gov/iad/humanid/feret/feret_master.html.

[9] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. pages 84–91, Jun 1994.

[10] Matthew A. Turk and Alex P. Pentland. Face recognition using eigenfaces. In *Proceedings of Computer Vision and Pattern Recognition*, pages 586–591, 1991.

[11] Thomas Vetter and Tomaso Poggio. Linear object classes and image synthesis from a single example image. *IEEE Transaction in Pattern Analysis and Machine Intelligence*, 19(7):733–742, July 1997.

[12] Laurenz Wiskott, Jean-Marc Fellous, Norbert Kruger, and Christoph von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.

[13] Ming-Hsuan Yang. Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. In *FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, page 215, Washington, DC, USA, 2002. IEEE Computer Society.

[14] Changshui Zhang, Jun Wang, Nanyuan Zhao, and David Zhang. Reconstruction and analysis of multi-pose face images based on nonlinear dimensionality reduction. *Pattern Recognition*, (37):325–336, 2004.

[15] Shaohua Kevin Zhou and Rama Chellappa. Illuminating light field: Image-based face recognition across illuminations and poses. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR04)*, 2004.