Synthesized GMM Free-parts Based Face Representation for Pose Mismatch Reduction in Face Verification

Simon Lucey
Advanced Multimedia Processing Laboratory
Department of Electrical and Computer Engineering
Carnegie Mellon University, Pittsburgh PA 15213, USA
slucey@ieee.org

Conrad Sanderson
CRC for Sensor Signal and Information Processing
Department of Electrical and Electronic Engineering
University of Adelaide, SA 5005, Australia
conradsand@ieee.org

Abstract

Performance of face verification systems can be adversely affected by mismatches between training and test poses, especially when only one pose is available for training. Compared to holistic/monolithic representations, we show that a "free-parts" representation of the face is less affected by pose changes, due to: a) some patches of a subject's face retaining similar appearance across a number of different poses, and b) those patches being able to freely move position across different poses. Furthermore, we propose that this mismatch can be reduced further by synthesizing the statistical model of a subject's "free-parts" representation for a set of poses for which there are no gallery observations. The synthesis is accomplished by first learning how a model for a generic frontal face transforms to represent a generic face at a particular non-frontal pose. The learned transformation is then applied to each subject's frontal model to synthesize a non-frontal model. The original and synthesized models are then concatenated in order to automatically handle multiple poses.

1. Introduction

Pose mismatch between a client's gallery and probe images is a very important problem in automatic face recognition at the moment. The pose mismatch problem can occur in applications such as person spotting in surveillance videos (e.g. at an airport). Generally the pose of faces in surveillance videos is uncontrolled, and there may be only one reference image (e.g. a passport photograph) for the person to be spotted.

Considerable work has already been performed with monolithic face representations, for automatic face recognition, in the presence of pose mismatch. Most notably techniques like Tensorfaces [1], Eigen-light fields [2] and Fisherfaces [3] have been employed with varying degrees of success. The term *monolithic* is employed in this paper to describe the holistic vectorized representation of the face based purely on pixel values within an image array, which can be associated with the holistic mechanism used in a human face recognition system [4].

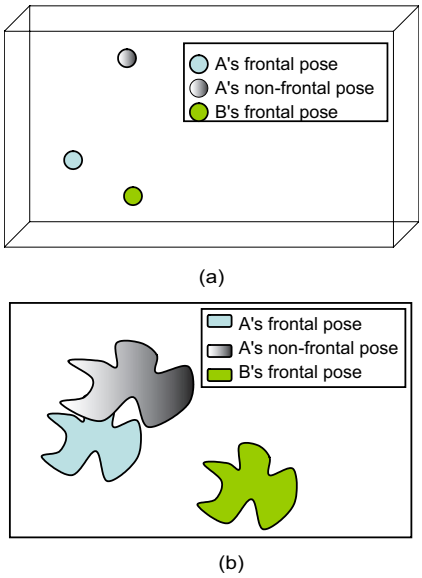

Figure 1: Depiction of our hypothesis on the benefit of matching distributions (free-parts representation) rather than points (monolithic representation) when dealing with pose variation. (a) Depicts an example of the unwanted scenario where the distance to Subject B's point from Subject A's point will be less than the distance to Subject A's non-frontal point. (b) Depicts the desirable example of where the divergence between Subject B's distribution and Subject A's distribution will be more than the divergence of Subject A's frontal and non-frontal distributions.

Our work is primarily motivated by the hypothesis that monolithic "point" based face representations of the face are much more prone to pose variation than parts "distribution" based representations. A graphical depiction of this hypothesis can be seen in Figure 1. We base this hypothesis on the assumption that the mismatch in appearance between viewpoints of the same subject is not homogeneous across all patches of a face image. The employment of

parts representations for object/face detection has recently gained much attention and success in machine vision literature [5, 6, 7]. The term *parts* is used to denote a representation of the face that can be considered as an ensemble of image patches of the image array; this representation can be considered similar to the component parts mechanism seen in a human face recognition system [4]. Parts representations have an advantage over monolithic representations in that they are able to assume varying dependencies between other patches within an image [5]. For all the work presented in this paper we shall assume minimal dependence between patches in an image. We make a further distinction in this paper between *free-parts* and *rigid-parts* representations.

Free-parts representations employ a strategy of where the position/structure of patches within the image can be relaxed so that these patches are able, to varying extents, "freely" move. Recently techniques that employ this representation have achieved good performance in frontal viewpoint face recognition tasks [8, 9, 10]. Generative models have been used in this previous work to model these free-parts face distributions such as pseudo 2-D hidden Markov models (HMMs) [10, 11] and Gaussian mixture models (GMMs) [8, 9]. GMMs can be thought of as a special subset of HMMs where no positional constraints are placed on the patch observations whatsoever; this is a highly desirable characteristic when trying to verify clients across pose, as patch positions can vary considerably across viewpoints.

Rigid-parts representations employ a strategy where the weight/contribution of each patch to the recognition process is not homogenous but the position/structure is preserved. There has already been some preliminary work by Kanade and Yamada [12] demonstrating the benefit of a rigid-parts representation when attempting to recognize faces across pose. However, these representations are still "point" based with each patch/point existing in a separate feature space. From this perspective free-parts representations are dissimilar to monolithic and rigid-parts representations as they generate many feature observations all existing within the same feature space. The central focus of our work is to demonstrate that there is considerable benefit in entertaining a "distribution" style representation of the face when there is a mismatch in viewpoint. Specific comparisons between free-parts and rigid-parts strategies will not be entertained in this paper.

In this paper we will demonstrate that the defining factor in deciding which face representation (i.e. monolithic or free-parts) and which learning algorithm to employ is the variability available in the development set. We define the development set as the set of observations used to obtain any data-dependent aspects of the verification algorithm (e.g. Eigenface vectors, world models, etc.), but does *not* provide any client specific information like those found in the gallery and probe sets.

Throughout this correspondence we make the following novel contributions to pose mismatch face verification. First, we demonstrate that free-parts representations, in the presence of pose mismatch, are inherently superior to monolithic representations when there is minimal pose variation available in the development set. Second, we propose a model synthesis approach that is able to make an estimate of a client's GMM for an unseen pose based on the client's GMM for a seen pose and prior knowledge obtained from the development set. We also provide evidence to suggest that when suitable pose variation does exist in the development set, the synthesis of unseen client pose GMMs is still non-trivial. We refer to this dilemma as the "model correspondence problem". Finally, we demonstrate that the effects of the correspondence problem can be reduced by employing a constrained version of relevance adaptation (RA) [8] on the GMM, which we refer to as "modified" RA.

2. Free-parts Representations

Learning the face as a distribution (i.e. many observations), as opposed to a single observation, has many appealing properties for face classification tasks. First, the many observations (representing the face) can exist in a low dimensional space circumventing problems associated with the "curse of dimensionality" [13] when training a classifier with high dimensional observations. Second, by representing a face with many observation points one naturally has more observations (of a lower dimensionality) to aid in the estimation of a classifier's parameters. Through the use of GMMs to model the face distribution, it has been shown [8, 9] that good verification performance can be attained by throwing away most position/structure information. We refer to this type of face model as a free-parts GMM (FP-GMM). In this subsection we briefly explain what features we use to estimate the FP-GMM, how it is estimated and how we evaluate the GMM during verification.

2.1. Free-parts GMMs

To estimate or evaluate a FP-GMM for a subject, the subject's geometrically and statistically normalized images are first decomposed into 16×16 pixel image patches with a 75% overlap between horizontally and vertically adjacent patches. Each image patch has a 2D-DCT applied to it in order to compact the 256 elements into a feature vector $\mathbf{o}$ of dimensionality D. Based on preliminary experiments, we have chosen $D = 35$. Additional information about the generation of the feature representations can be obtained from [8, 9].

A GMM models the probability distribution of a D dimensional random variable $\mathbf{o}$ as the sum of M multivariate Gaussian functions,

$$f(\mathbf{o}|\boldsymbol{\lambda}) = \sum_{m=1}^{M} w_m \mathcal{N}(\mathbf{o}; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \quad (1)$$

where $\mathcal{N}(\mathbf{o}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the evaluation of a normal distribution for observation $\mathbf{o}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The weighting of each mixture component is denoted by w_m and must sum to unity across all components. In our work the covariance matrices in $\boldsymbol{\lambda}$ are assumed to be diagonal such that $\boldsymbol{\Sigma} = \text{diag}\{\boldsymbol{\sigma}\}$, as substantial benefit can be attained by reducing the number of parameters that need to be estimated.

Given a world model $\boldsymbol{\lambda}_{\mathtt{w}} = \{w_{\mathtt{w}m}, \boldsymbol{\mu}_{\mathtt{w}_m}, \boldsymbol{\Sigma}_{\mathtt{w}_m}\}_{m=1}^{M}$ and training observations from a particular client, $\mathbf{O} = \{\mathbf{o}_1, \cdots, \mathbf{o}_R\}$, the GMM parameters for that client are estimated through relevance adaptation (RA) [8].

The world model is simply a single model trained from a large number of subject faces representative of the general population. The world model's parameters are estimated using the Expectation Maximization (EM) algorithm [14], configured to maximize the likelihood of training data. RA is an instance of the EM algorithm configured for maximum *a posteriori* (MAP) estimation, rather than simply maximum likelihood (ML). It has been noted that great benefit can be obtained in terms of estimating high performance robust FP-GMMs by employing RA when only small amounts of client specific observations exist (e.g. a single enrollment image). Using RA, parameters for client $\mathtt{c}$ are obtained using the following update equations:

$$w_{\mathtt{c}_m} = \beta \left[(1 - \alpha_m^w) w_{\mathtt{w}_m} + \alpha_m^w \frac{\sum_{r=1}^{R} \gamma_m(\mathbf{o}_r)}{\sum_{m=1}^{M} \sum_{r=1}^{R} \gamma_m(\mathbf{o}_r)} \right] \quad (2)$$

$$\boldsymbol{\mu}_{\mathtt{c}_m} = (1 - \alpha_m^{\mu}) \boldsymbol{\mu}_{\mathtt{w}_m} + \alpha_m^{\mu} \frac{\sum_{r=1}^{R} \gamma_m(\mathbf{o}_r) \mathbf{o}_r}{\sum_{r=1}^{R} \gamma_m(\mathbf{o}_r)} \quad (3)$$

$$\boldsymbol{\sigma}_{\mathtt{c}_m} = (1 - \alpha_m^{\sigma}) \left(\boldsymbol{\sigma}_{\mathtt{w}_m} + \boldsymbol{\mu}_{\mathtt{w}_m}^2 \right) + \alpha_m^{\sigma} \frac{\sum_{r=1}^{R} \gamma_m(\mathbf{o}_r) \mathbf{o}_r^2}{\sum_{r=1}^{R} \gamma_m(\mathbf{o}_r)} - \boldsymbol{\mu}_{\mathtt{c}_m}^2 \quad (4)$$

where $\gamma_m(\mathbf{o})$ is the occupation probability for component m, $\boldsymbol{\mu}^2$ indicates that each element in $\boldsymbol{\mu}$ is squared, and α_m^{ρ} is a weight used to tune the relative importance of the prior; it is defined as:

$$\alpha_m^{\rho} = \frac{\sum_{r=1}^{R} \gamma_m(\mathbf{o}_r)}{\tau^{\rho} + \sum_{r=1}^{R} \gamma_m(\mathbf{o}_r)} \quad (5)$$

where τ^{ρ} is a *relevance* factor. The above definition of α_m^{ρ} can limit the adaptation to only the Gaussians for which there is sufficient data. We have found effective performance can be attained by using a single relevance factor ($\tau = \tau^w = \tau^{\mu} = \tau^{\sigma}$). Based on empirical evaluation on many data sets, we have chosen $\tau = 10$. The scale factor, β, in Equation 2 is computed to ensure that all the adapted component weights sum to unity. The adaptation procedure is iterative, thus an initial client model is required. This is accomplished by copying the world model.

In RA, the distributions are estimated by finding and using observations that aid in discriminating client models from the world model. As such, the distributions should not be considered as generative distributions (i.e. distributions that can be used for producing synthetic observations representative of a particular client). In this sense the GMM based classifier, trained via RA, is inherently discriminative and is able to obtain good classification performance with sparse amounts of training data. Additional information on RA can be found in [8].

2.2. Evaluating a FP-GMM

To evaluate a sequence of observations, generated from a claimant's probe image, we obtain the average log-likelihood,

$$\mathcal{L}(\mathbf{O}|\boldsymbol{\lambda}_c) = \frac{1}{R} \sum_{r=1}^{R} \log f(\mathbf{o}_r|\boldsymbol{\lambda}_c) \quad (6)$$

Given the average log-likelihood, for the client and world models, one can then calculate the log-likelihood ratio,

$$\Lambda(\mathbf{O}) = \mathcal{L}(\mathbf{O}|\boldsymbol{\lambda}_c) - \mathcal{L}(\mathbf{O}|\boldsymbol{\lambda}_w) \quad (7)$$

For our work we found good performance across pose could be attained if we employed GMMs with 32 components.

3. Model Synthesis

We propose a model synthesis approach that is able to estimate a client's FP-GMM for an unseen pose, based on the client's FP-GMM for a seen pose and prior knowledge obtained from the development set. In the proposed model synthesis approach, prior information is used to construct world face models for different views through a *modified* form of the RA algorithm (described in Section 3.1). The synthesis is accomplished by first learning how the frontal world model differs from a non-frontal world model. The differences between the models are comprised of the differences between the means in the corresponding Gaussians (i.e. how the means have moved) and the differences between the covariance matrices (i.e. how the diagonal entries in the covariance matrices have scaled). Weights are not considered as empirical observations show that the differences are almost entirely reflected in the means and covariances.

Let us denote the frontal world model as $\boldsymbol{\lambda}_w^{0^o}$ and the non-frontal world model for angle Θ as $\boldsymbol{\lambda}_w^{\Theta}$. The set of parameters which describes the differences is formally defined as:

$$\Psi = \{ \boldsymbol{\Delta}_m, \boldsymbol{s}_m \}_{m=1}^{M} \quad (8)$$

The parameters of the above set are in turn defined as:

$$\boldsymbol{\Delta} = \boldsymbol{\mu}_w^{\Theta} - \boldsymbol{\mu}_w^{0^o} \quad (9)$$

$$\boldsymbol{s}^T = [\, s_d \,]_{d=1}^{D} = \left[\boldsymbol{\sigma}_{w,(d)}^{\Theta} / \boldsymbol{\sigma}_{w,(d)}^{0^o} \right]_{d=1}^{D} \quad (10)$$

where the subscript denoting the mixture component index m has been dropped to improve clarity. The notation $\boldsymbol{\sigma}_{\mathtt{w},(d)}$ denotes element d of covariance vector $\boldsymbol{\sigma}_{\mathtt{w}}$.

Since the two world models are a good representation of a general face at two different views, and each frontal client model is derived from the frontal world model, we conjecture that we can apply the above differences to client $\mathtt{c}$'s frontal model in order to synthesize a model for angle Θ. Formally, the parameters for each Gaussian in client $\mathtt{c}$'s Θ model are found using:

$$w_{\mathtt{c}}^{\Theta} = w_{\mathtt{c}}^{0^o} \quad (11)$$

$$\boldsymbol{\mu}_{\mathtt{c}}^{\Theta} = \boldsymbol{\mu}_{\mathtt{c}}^{0^o} + \boldsymbol{\Delta} \quad (12)$$

$$\boldsymbol{\sigma}_{\mathtt{c}}^{\Theta} = \boldsymbol{s} \star \boldsymbol{\sigma}_{\mathtt{c}}^{0^o} \quad (13)$$

where $\star$ indicates element by element multiplication.

3.1 Model Correspondence Problem

The synthesis technique described above pre-supposes that there is a correspondence between Gaussian mixture components of the client's frontal model, the frontal and non-frontal world models; we define correspondence as each Gaussian describing the *same aspects* of the face. However, under the training paradigm of RA, where an existing GMM is iteratively adapted with new data, there is no explicit guarantee that Gaussians in the resulting model will correspond to the Gaussians in the original model. This problem arises from the nature of the EM algorithm, which in the case of GMM parameter estimation can be considered as a form of unsupervised soft clustering.

To address the correspondence issue, we propose to modify RA and the EM algorithm upon which it is based. Let us first define a "parent model" as the model to be adapted and a "child model" as the model that resulted from adapting a "parent model"; in a similar vein, let us define a "parent Gaussian" as a Gaussian from the "parent model" and a "child Gaussian" as the Gaussian that resulted from a particular "parent Gaussian" through the process of adaptation. We wish to prevent the child Gaussians modeling different aspects of the face than their parents by inhibiting each child Gaussian from moving too far away from its parent. We will assume that when a child Gaussian is closer to some other child's parent than its own parent, it has moved too far.

Let us define the distance between two Gaussians as the Mahalanobis distance [13] between their means:

$$\mathcal{M}(\boldsymbol{\mu}_a, \boldsymbol{\mu}_b) = (\boldsymbol{\mu}_a - \boldsymbol{\mu}_b)^T \boldsymbol{\Sigma}_{\mathtt{all}}^{-1} (\boldsymbol{\mu}_a - \boldsymbol{\mu}_b) \quad (14)$$

where $\boldsymbol{\Sigma}_{\mathtt{all}} = \mathrm{diag}(\boldsymbol{\sigma}_{\mathtt{all}})$ is the overall covariance matrix of the parent world model. It can be shown that $\boldsymbol{\sigma}_{\mathtt{all}}$ is found using:

$$\boldsymbol{\sigma}_{\mathtt{all}} = -\boldsymbol{\mu}_{\mathtt{all}}^2 + \sum_{m=1}^{M} w_m (\boldsymbol{\sigma}_m + \boldsymbol{\mu}_m^2) \quad (15)$$

where $\boldsymbol{\mu}_{\mathtt{all}} = \sum_{m=1}^{M} w_m \boldsymbol{\mu}_m$. Note that we have omitted the subscript $\mathtt{w}$ from the world model's parameters for clarity.

The RA algorithm is modified by introducing an early stopping criterion. At the end of each iteration a check is made to see if any child Gaussian is too far away from its parent. If this has occurred, the parameters from the last iteration are restored and the RA process is deemed to have converged. The check is enabled from the second iteration onwards in order to ensure the child model is different from the parent model.

4. Monolithic representations

It is outside the scope of this paper to perform a large scale evaluation of all possible monolithic approaches. Instead we will be taking a sample of techniques that are representative of current paradigms in pose robust face recognition. Specifically, we will be considering the Eigenface algorithm [15] as a baseline due is ubiquitous nature in face recognition literature. The Fisherface algorithm [16] is also considered as a baseline due to its simplicity and high performance in recent evaluations [17, 18, 19]. Finally, the Eigenlight-fields technique will be used as a baseline due to its specificity to pose and its similar nature to other popular approaches such as Tensorfaces [1].

4.1. Eigen- and Fisher-faces

Eigen- and Fisher-face approaches have been around for quite some time and have enjoyed much success in frontal face recognition. In this paper we will be evaluating a specific type of Eigen- and Fisher-face strategy. The first, which will be referred to as MON-PCA, is the baseline Eigenface [15] technique which employs principal component analysis (PCA) to generate a subspace preserving the $K = 89$ most energy preserving modes. The whitened cosine distance (i.e. the cosine distance between two observations after performing the whitening transform [13] on both of them) is then employed to gain a measure of similarity between the gallery and probe observation vectors which result after mapping the original pixel images into the PCA generated subspace. The second technique, which we shall refer to as MON-LDA, is a variant on the Fisherface [16] technique which employs linear discriminant analysis (LDA), after an initial PCA stage, to generate a subspace preserving the $K = 89$ most discriminant modes. As suggested by [17, 18, 19] good performance can be attained if we employ the cosine distance to gain a measure of similarity.

4.2. Eigen-light Field Approach

Eigen-light fields were proposed by Gross *et al.* [2] as a technique for learning the dependencies that exist between monolithic representations of the face from different view

points. In their paper Gross *et al.* argue that a face's light field is an ideal representation to perform face recognition under varying pose as the representation naturally encompasses all view points. A face was assumed to stem from only a finite set of poses $1, 2, \ldots, P$. In their work a light field was represented as the concatenation of the vectorized view point images. Canonical PCA was then applied, in a similar manner to the Eigenface approach, to build a compact representation of a subject's light field. The subspace which results from the PCA process, which the eigen-light field vectors span, is created by preserving the $K = 89$ most energy preserving modes. In practice however, one rarely has all possible view points to construct a complete light field. In fact, it is quite common to only have a single gallery view point. In this common scenario a least squares approximation of the compact representation can be made from the incomplete representation [2]. Gross et al. demonstrated that this approach performed well in comparison to the Eigenface algorithm and a commercial system. We employed the cosine distance to generate match-scores for verification. Throughout the experimental portion of this paper we shall refer to this specific technique as LF-PCA.

5. Face Database and Normalization

Experiments were performed on a subset of the FERET database [20], specifically images stemming from the *ba*, *bb*, *bc*, *bd*, *be*, *bf*, *bg*, *bh*, and *bi* subsets; which approximately refer to rotation's about the vertical axis of 0^o, $+60^o$, $+40^o$, $+25^o$, $+15^o$, -15^o, -25^o, -40^o, -60^o respectively. The database contains 200 subjects which were randomly divided into an evaluation and development set both containing 90 subjects. The remaining 20 subjects were used as an imposter set for our verification experiments. The development set is used to obtain any data-dependent aspects of the verification system (e.g. subspace, world models etc.). The evaluation and imposter set is where the performance rates for the verification system are obtained.

Traditionally, before performing the act of face recognition, some sort of geometric pre-processing has to go on to remove variations in the face due to rotation and scale. The distance d_{eye} and angle θ_{eye} between the eyes has long been regarded as an accurate measure of scale and rotation in a face. However, this type of geometric normalization, based purely on the eye position, becomes problematic when faced with depth pose rotation. An example of this problem can be seen in row 1 of Figure 2.

Essentially this type of normalization becomes more and more problematic, in terms of stretching the image along the y-axis thus changing the aspect ratio of the cropped face image. An obvious way to remedy this situation is to employ a distance on the face that gives additional vertical information to circumvent this aspect ratio change. The distance from the eye line to the nose tip vertically, d_{nose}, is an obvious choice. One can see that the aspect ratio problem is considerably diminished in row 2 of Figure 2. One can also see however that the mouth still remains a problem as it changes dramatically in appearance across pose. In this paper we additionally address this problem by cropping out the mouth before passing the image to a verification algorithm. An example of these final cropped images can be seen in row 3 of Figure 2. The final geometrically normalized cropped faces formed an 98×115 array of pixels.

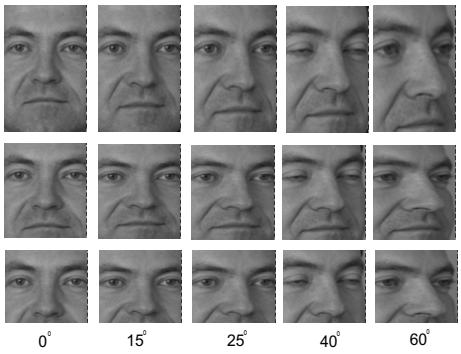

Figure 2: Problem of aspect-ratio across pose when using d_{eye}, left horizontal sweep shown in row 1. One can see, in row 2, that employing the d_{nose} distance greatly alleviates this effect. One can further see in row 3 that cropping the bottom of the mouth increases the invariance to pose.

6. Face verification task

The face verification task is the binary process of accepting or rejecting the identity claim (i.e. the log-likelihood ratio or cosine distance match-score from the free-parts and monolithic recognizers respectively) made by a subject under test. A threshold Th needs to be found so as to make the decision. Face verification performance is evaluated in terms of two types of error: a) being false rejection (FR) error, where a true client is rejected against their own claim, and b) false acceptance (FA) errors, where an impostor is accepted as the falsely claimed subject. The FA and FR errors increase or decrease in contrast to each other based on the decision threshold Th set within the system. A simple measure for overall performance of a verification system is found by determining the equal error rate (EER) for the system, where FA = FR.

7. Results and Discussion

We have elected to present results for two particular cases in order to demonstrate the full benefit of a free-parts representation for pose mismatched face recognition.

Pose	MON-PCA$^{\circledast}$	MON-PCA
-60	27.77	21.36
-40	20.00	15.56
-25	12.22	8.89
-15	7.78	7.70
15	6.61	4.44
25	10.00	10.00
40	18.95	16.67
60	24.44	18.89
Average	15.97	12.94

Table 1: Results illustrating the importance of enrolling a MON-PCA representation with eigenface vectors that have been estimated with a development set containing both frontal & non-frontal poses, relative to those employing a development set that contains frontal only observations. Models estimated with the frontal-only data are denoted with $\circledast$. Results are in terms of EER (%).

Pose	FP-GMM$^{\circledast}$	FP-GMM
-60	19.58	26.10
-40	10.33	10.58
-25	5.33	5.35
-15	2.81	2.65
15	3.14	4.22
25	8.08	8.14
40	15.19	24.17
60	23.78	37.08
Average	11.03	14.79

Table 2: Results illustrating the importance of training FP-GMMs with a world model that is estimated with development set containing *only* frontal poses relative to those employing a development set that contains frontal and non-frontal observations. Models estimated with the frontal-only world model are denoted with $\circledast$. Results are in terms of EER (%).

1. For the case where a pose mismatch exists between the probe set and the gallery & development set. In this circumstance the development set contains only the poses present in the gallery set (i.e. frontal).
2. For the case where a pose mismatch exists only between the probe and gallery sets. In this circumstance the development set contains all the poses that will be present in the probe and gallery sets (i.e. frontal and non-frontal).

It is obvious for some techniques such as Fisherfaces (MON-LDA) and Eigen-light Fields (LF-PCA) that the employment of frontal and non-frontal development observations is necessary as they are intrinsic to their framework. These techniques can only be analyzed under the situation described in Case 2. However, techniques like Eigenfaces (MON-PCA) and Free-parts GMMs (FP-GMM) can actually be constructed employing either frontal only or frontal & non-frontal development observations. The motivation for this analysis is to: first, investigate whether the employment of additional pose information in the development set is necessary for good performance in the presence of pose mismatch; and second, is there a difference in how each representation (i.e. monolithic or free-parts) performs depending on the variability available in the development set.

7.1. Case I

Table 1[1] depicts results for the monolithic MON-PCA technique using frontal only and frontal & non-frontal development sets in the creation of their eigenface vectors.

One can see in Table 1 that the employment of a development set without non-frontal observations, as denoted by the absence of a $\circledast$, in the creation of eigenface vectors has a catastrophic affect on performance when the probe images are non-frontal. The results demonstrate that tuning the eigenface vectors to frontal only views seriously affects performance, as they are unable to adequately represent the non-frontal poses. This is inline with our intuitive thoughts of the benefits of providing subject independent prior knowledge, during training, about all possible variations that will be encountered in testing.

The results in Table 2 however depict an opposite result for the FP-GMM algorithm, employing the free-parts representation, with respect to what pose variation should be present in the development set. Table 2 demonstrates there is actually substantial benefit in employing a world model that has been estimated from frontal only development observations as denoted by the $\circledast$. A satisfactory explanation of these results can be formed if one takes into account the nature of how the development set observations are being employed by the FP-GMM.

First, the FP-GMM algorithm employs the development observations to create a background class (i.e. world model) for the enrolled client to discriminate against during the estimation of the GMM. Depending on what type of variation is contained in the development set will dictate what the GMM will be discriminating against. For example, if the client set only contains observations from a frontal pose of the client but the development set contains observations from many subjects across many poses the resultant GMM will be discriminatory against the claimant's identity *and* pose. However, if the development set contained only observations from a frontal pose across many subjects then the resultant GMM will be discriminatory against the claimant's identity *not* pose. How habile the resultant GMM is to different poses is dependent on how well that representation generalizes across pose. The MON-PCA algorithm employs a match-score metric, namely the whitened cosine distance, that is dependent of the development set. Unlike the FP-GMM log-likelihood ratio match-score metric however, the whitening process of the whitened cosine distance

[1]Note: throughout the entire results section all techniques that begin with a MON-, LF- and FP- label refer to monolithic, light field and free-parts feature representations respectively. The subsequent PCA, LDA or GMM label refers to which subspace or classifier they employed to generate the match-score.

Pose	Synthesized using std. RA	Synthesized using mod. RA	Concatenation synth. mod. RA
-60	+0.84	-3.05	-2.75
-40	-0.64	-2.19	-2.61
-25	-0.36	-0.41	-0.59
-15	+1.47	+0.30	-0.03
15	+0.17	-0.36	-0.20
25	-1.16	-1.77	-1.80
40	-2.08	-5.88	-5.30
60	-6.00	-8.25	-5.78
Average	-0.97	-2.70	-2.23

Table 3: Results for synthesized FP-GMMs in terms of their relative difference (+/-) in EER (%) to the standard frontal model results listed in Table 2. Column 1 denotes results for synthesized models employing standard RA. Column 2 denotes results for synthesized models employing modified RA. Column 3 denotes results for the synthesized models estimated using modified RA when the pose of the probe image is unknown, and the pose models are concatenated together.

is dependent on seeing pose variation in the development set (in terms of the eigenvalues of the PCA process) that will be seen in the probe set.

Second, the FP-GMM process employs a data-independent feature extraction process (i.e. 2D-DCT) which is in no way dependent on the development set. This is a highly advantageous characteristic in comparison to the MON-PCA algorithm. The MON-PCA approach employs a feature extraction process that is extremely dependent on the development set. If variations, whether they be pose or subject, are present in the probe set that are not seen in the development set then the ability to obtain features that are able to represent these variations is seriously affected. For the rest of the results in this correspondence the FP-GMMs employing frontal only development observations shall be referred to as *standard frontal* FP-GMMs.

7.2. Case II

Table 3 shows the results for client models synthesized for a specific angle, while using standard and modified RA. The results for standard RA show only minor improvements in performance, while for modified RA there is a considerably greater improvement. These results thus support the use of modified RA in order to address and show evidence for the existence of the model correspondence problem.

If synthesized models were to be used in a real-life application, then a pose detection step would be necessary to select the most appropriate model. In order to mitigate the need for pose detection, we have investigated the use of FP-GMMs which represent a face at many poses. Such FP-GMMs were obtained by concatenating each client's frontal FP-GMM with FP-GMMs synthesized for specific angles. The frontal world model was also concatenated with non-frontal world models. Since each FP-GMM had 32 components, each resulting concatenated FP-GMM had $32 \times 9 = 288$ components. Results in Table 3 show that for most angles the concatenated models obtain only a small reduction in performance, when compared to models synthesized for specific angles.

In Figure 3 one can see the final breakdown of performance between leading monolithic and free-parts face verification approaches that are able to make use of pose variation in the development set. One can see the FP-GMM approach, which in this instance is employing model synthesis and modified RA, either outperforms or is approximately equal to the performance of both the LF-PCA and MON-LDA approaches except at the extreme view points of +/- 60^o. In this instance the both monolithic approaches outperform our free-parts approach with the Fisherface based MON-LDA obtaining best performance. A partial explanation for this result could be found in the strong assumptions (i.e. view point changes result in a global mean shift and variance scaling across subjects for each Gaussian in the FP-GMM) we have made during the model synthesis process start to break down at extreme view points. The Eigen-light field based LF-PCA approach fares the worst on average across poses although performance is consistent across all poses. Future work shall try to relax some of these assumptions to improve performance at these extreme view points. Investigating alternate patch sizes and relevance factors may also reduce the severity of the discrepancy between monolithic and free-parts performance at these larger non-frontal viewpoints. From these results one can see that there is clear benefit in exploring both free-parts and monolithic representations when developing a face verification system that is robust to pose mismatches; provided there is ample pose variation in the development set.

8. Summary and Conclusions

The verification results presented here convincingly demonstrate that a free-parts representation of the face is beneficial in the presence of a pose mismatch. In our work we were able to demonstrate the habile behavior of FP-GMMs to pose mismatch, in comparison to monolithic approaches, when the development set does not contain any pose variation. This result is significant as it demonstrates that free-parts representations can be used with some success in pose circumstances that have not been seen in the development set.

We were also able to demonstrate that improved performance can be achieved with FP-GMMs when there is pose variation present in the development set, even when the pose of the probe image was unknown. This improved performance was attained through the synthesis of unseen pose models through transformations learned from the development set along with the employment of the modified RA.

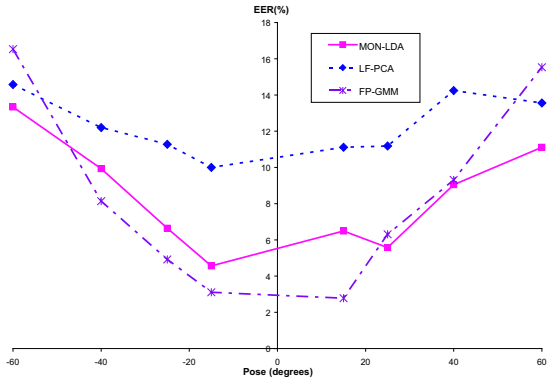

Figure 3: Comparison between leading monolithic (Eigen-light Fields (LF-PCA) and Fisherfaces (MON-LDA) approaches) and free-parts (FP-GMMs using synthesis and modified RA) algorithms that make use of pose variation in the development set. Results demonstrate improved or equivalent performance for our free-parts algorithm over leading monolithic algorithms in all but the most extreme view point(+/- 60^o).

References

[1] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: TensorFaces," in *European Conference on Computer Vision (ECCV)*, vol. 2350 of *Lecture Notes in Computer Science*, (Berlin), pp. 447–460, Springer-Verlag, 2002.

[2] R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and light-fields," *IEEE Trans. PAMI*, vol. 26, pp. 449–465, April 2004.

[3] H. Lee and D. Kim, "Pose invariant face recognition using linear pose transformation in feature space," in *European Conference on Computer Vision (ECCV)*, 2004.

[4] J. W. Tanaka and M. J. Farah, "The holistic representation of faces," in *Perception of Faces, Objects, and Scenes* (M. A. Peterson and G. Rhodes, eds.), ch. 2, pp. 53–74, Oxford University Press, Inc., 2003.

[5] H. Schneiderman and T. Kanade, "A histogram-based method for detection of faces and cars," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 504–507, September 2000.

[6] M. Weber, M. Welling, and P. Perona, "Unsupervised learning of models for recognition," in *European Conference on Computer Vision (ECCV)*, pp. 18–32, 2000.

[7] M. Weber, M. Welling, and P. Perona, "Towards automatic discovery of object categories," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 101–108, June 2000.

[8] S. Lucey and T. Chen, "A GMM parts based face representation for improved verification through relevance adaptation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. II, (Washington D.C.), pp. 855–861, June 2004.

[9] C. Sanderson and K. Paliwal, "Fast features for face authentication under illumination direction changes," *Pattern Recognition Letters*, vol. 24, no. 14, pp. 2409–2419, 2003.

[10] S. Eickeler, S. Muller, and S. Rigoll, "Recognition of JPEG compressed face images based on statistical methods," *Image and Vision Computing*, vol. 18, no. 4, pp. 279–287, 2000.

[11] F. Cardinaux, C. Sanderson, and S. Bengio, "Face Verification Using Adapted Generative Models," in *Proc. 6th IEEE Int. Conf. Automatic Face and Gesture Recognition (AFGR)*, (Seoul), pp. 825–830, 2004.

[12] T. Kanade and A. Yamada, "Multi-subregion based probabilistic approach toward pose-invariant face recognition," in *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, (Kobe, Japan), pp. 954–958, July 2003.

[13] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York, NY, USA: John Wiley and Sons, Inc., 2nd ed., 2001.

[14] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Royal Statistical Society*, vol. 39, pp. 1–38, 1977.

[15] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, 1991.

[16] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. PAMI*, vol. 19, no. 7, pp. 711–720, 1997.

[17] P. Navarrete and J. Ruiz-del-Solar, "Analysis and comparison of eigenspace-based face recognition approaches," *Int. Journal of Pattern Recognition and Artificial Intelligence*, vol. 16, no. 7, pp. 817–830, 2002.

[18] J. Ruiz-del-Solar and P. Navarrete, "Towards a generalized eigenspace-based face recognition framework," in *4th Int. Workshop on Statistical Techniques in Pattern Recognition*, (Windsor, Canada), August 2002.

[19] M. Sadeghi, J. Kittler, A. Kostin, and K. Messer, "A comparative study of automatic face verification algorithms on the BANCA database," in *AVBPA*, pp. 35–43, 2003.

[20] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. PAMI*, vol. 10, no. 22, pp. 1090–1104, 2000.