Learning Patch Dependencies for Improved Pose Mismatched Face Verification

Simon Lucey   Tsuhan Chen

Carnegie Mellon University

slucey@ieee.org, tsuhan@cmu.edu

Abstract

Most pose robust face verification algorithms, which employ 2D appearance, rely heavily on statistics gathered from offline databases containing ample facial appearance variation across many views. Due to the high dimensionality of the face images being employed, the validity of the assumptions employed in obtaining these statistics are essential for good performance. In this paper we assess three common approaches in 2D appearance pose mismatched face recognition literature. In our experiments we demonstrate where these approaches work and fail. As a result of this analysis, we additionally propose a new algorithm that attempts to learn the statistical dependency between gallery patches (i.e. local regions of pixels) and the whole appearance of the probe image. We demonstrate improved performance over a number of leading 2D appearance face recognition algorithms.

1. Introduction

It is often tempting to look at the problem of face recognition[1], in the presence of pose mismatch, as a 3D computer vision problem. Pose variation, in terms of pixel appearance, is highly non-linear in 2D, but linear in 3D. Notable work, such as that conducted by Vetter and Blantz [1], has attempted to leverage this inherent 3D advantage for face recognition in the presence of pose mismatch. A drawback to this approach however, is the requirement for multiple gallery images or depth information of the subject. In many face verification applications, we have at most a single 2D gallery image. This will be assumed throughout the work presented in this paper.

For the task of 2D appearance face verification we want to obtain the likelihood functions,

$$p(\mathbf{x}^g, \mathbf{x}^p | \omega), \quad \omega \in \{\mathcal{C}, \mathcal{I}\} \qquad (1)$$

"Gallery Image"   "Probe Image"

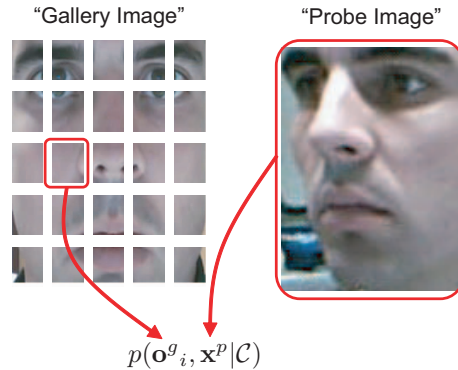$$p(\mathbf{o}^g{}_i, \mathbf{x}^p | \mathcal{C})$$

Figure 1. In this paper we demonstrate that good performance, which is robust to pose mismatch, can be obtained by modeling the marginal distribution of gallery patch appearance $\mathbf{o}^g$ at position i with the whole appearance of the probe image $\mathbf{x}^p$ (note we employ the notation $\mathbf{x}$ for representing the whole facial appearance, and $\mathbf{o}$ for representing patch appearance).

where ω refers to the classes where the gallery ($\mathbf{x}^g$) and probe ($\mathbf{x}^p$) images are similar ($\mathcal{C}$) and dissimilar ($\mathcal{I}$) in terms of subject identity. There is no need in this formulation for subject labels, as we assume there is only a single gallery and probe image per subject.

Employing these likelihoods one can then apply Bayes rule to obtain,

$$P(\mathcal{C}|\mathbf{x}^g, \mathbf{x}^p) = \frac{P(\mathcal{C})p(\mathbf{x}^g, \mathbf{x}^p|\mathcal{C})}{P(\mathcal{C})p(\mathbf{x}^g, \mathbf{x}^p|\mathcal{C}) + P(\mathcal{I})p(\mathbf{x}^g, \mathbf{x}^p|\mathcal{I})} \qquad (2)$$

where $P(\mathcal{C})$ and $P(\mathcal{I})$ are priors for the client and imposter classes respectively. The probability in Equation 2 can be employed as a match-score in the task of face verification[2].

A contribution is made in this paper towards improved face verification performance, in the presence of pose mis-

---

[1]Even though our paper is dealing solely with the task of face verification, we have seen that it is appropriate and consistent to included literature from the broader face recognition community. For the purposes of this paper we shall define recognition as encapsulating the tasks of verification and identification.

[2]We must note that in many instances in literature [2, 3, 4], authors do not adhere strictly to a Bayesian formulation in estimating the match-score, instead opting for non-probabilistic measures (e.g. cosine, Euclidean, Mahalanobis distances) for calculating similarity. However, a commonality exists amongst all these techniques in terms of how they gather statistics of facial appearance variation. Throughout this paper we will be using likelihood functions, as found in Equation 1, for expressing these differing methods for modeling gallery-probe facial appearance.

match, through a method for learning the dependencies between individual gallery patches and the whole appearance of the probe image (see Figure 1). This approach is advantageous over existing pose invariant algorithms [5, 6, 7] due to the ability of our approach to learn dependencies between the gallery and probe images at the patch-level; without having to assume a specific alignment between the gallery and probe images. Our experimental results, in Section 7, demonstrate improved performance over current algorithms in literature.

The central focus of this paper will be questioning a number of assumptions, made previously in literature [5, 6, 7], for the task of 2D appearance based face recognition in the presence of pose mismatch. Hitherto, many of these assumptions have not been analyzed in a comparative manner. Specifically, we will be investigating the following modeling approaches:

A1: Even though we are limited to a finite world set, we should attempt to model the *joint appearance* likelihood of the whole face $p(\mathbf{x}^g, \mathbf{x}^p|\omega)$ directly [6, 8].

A2: We should attempt to learn the likelihood function for the *differential appearance* $p(\mathbf{x}^g - \mathbf{x}^p|\omega)$ between the whole probe and gallery images, rather than the joint likelihood $p(\mathbf{x}^g, \mathbf{x}^p|\omega)$. This assumption [3, 4, 7] is motivated largely by the realization that, provided the gallery and probe images are appropriately aligned, the relative difference in appearance contains a lot of discriminatory information; with the added advantage of not having to model the degrees of freedom present in $p(\mathbf{x}^g, \mathbf{x}^p|\omega)$.

A3: Individual image patches within the face should be assumed to be *statistically independent* of one another. This strong assumption is motivated by the hypothesis [5] that modeling the appearance variation of the entire face, in the presence of varying pose, is too difficult. By treating the face as an ensemble of independent patches we can, to some extent, circumvent this problem by learning the local appearance variation of each local region (i.e. patch) in the face.

In this paper we investigate the merit of these three modeling approaches with respect to gaining a tractable estimation to the likelihoods functions in Equation 1. As a result of this analysis we derive a novel approach (see Figure 1), that attempts to leverage the beneficial aspects of these three approaches into a unifying framework. Improved verification results are reported, relative to existing algorithms, on the pose portion of the FERET database [9].

2. Offline Training

Face verification is a task of generalization, where one attempts to gain an adequate estimate of the likelihood func-
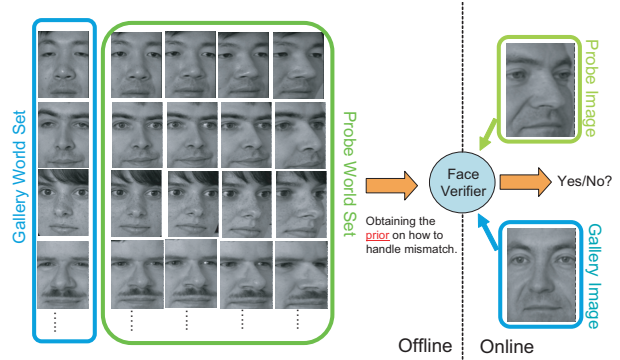

Figure 2. In any face recognition system there is an offline and online component. The offline component trains the face verifier on a *world* set, which is representative of the appearance variation we anticipate seeing in the gallery and probe images.

tion in Equation 1, from what we observe in a finite *world* set (see Figure 2). A high-level example of a face verifier is depicted in Figure 2 for the case of pose variation. In this example we depict how a face verifier relies on offline training from a world set. The world set contains a large number of subject faces representative of the population of subject faces expected during verification, but are usually independent, in terms of identity, to the subjects involved in the online verification process. In all the experiments within this paper we employ the world set for obtaining our statistics describing the client and imposter facial appearance classes.

3. Background

Techniques that decompose the face into an ensemble of salient patches/regions (e.g. eyes, nose, and mouth) have reported superior face recognition performance with respect to approaches that treat the face as a whole. Notable examples can be found in the work of Brunelli et al. [10], Moghaddam et al. [7] and Martinez [11]. By analyzing the face in terms of these separate salient patches/regions, good performance was noted in the presence of occlusion and expression mismatches [11]. Analyzing faces in terms of an ensemble of salient patches however, has not been successfully applied when there is pose mismatch. A reason for this poor performance may perhaps stem from the heuristic choice of salient patches, and that the discrimination provided by these heuristically chosen patches vary dramatically as a function of pose.

3.1. A1: Joint Appearance

Recently techniques, like the Eigen Light-Fields [6] and Tensorface [8] approaches, have found benefit in modeling the joint client likelihood function $p(\mathbf{x}^g, \mathbf{x}^p|\mathcal{C})$ directly in the presence of pose mismatch. The likelihood func-

tion $p(\mathbf{x}^g, \mathbf{x}^p|\mathcal{I})$ is typically omitted due to the complexity associated with its estimation. Due to the large dimensionality of the whole face $\mathbf{x}$, subspace approaches based on principal component analysis (PCA) must be employed to approximate the likelihood $p(\mathbf{x}^g, \mathbf{x}^p|\mathcal{C})$ from the finite world set. To match a known gallery image $\mathbf{x}^g$ against an unknown probe image $\mathbf{x}^p$ the following approximation is then made (note: the italicized vector x is a variable),

$$\mathbf{x}^{g*} \approx \int_x x\, p(x, \mathbf{x}^p|\mathcal{C}) dx \qquad (3)$$

In practice this approximation can be obtained simply, through a psuedo-inverse operation on the probe specific portion of the eigenvectors describing the joint appearance of the gallery and probe client images. The vectors $\mathbf{x}^g$ and $\mathbf{x}^{g*}$ are then matched against one another using a canonical distance measure. In our experiments we chose to use the cosine distance due to its good performance.

3.2. A2: Differential Appearance

As previously mentioned, a number of approaches have been employed in literature in order to estimate the 2D appearance likelihoods in Equation 1. The most well known of these has been the intra-personal (i.e. client) and extra-personal (i.e. imposter) approach of Moghaddam and Pentland [7]. In this approach the authors attempt to model the *differential appearance* between probe and gallery images $\mathbf{x}^p$ and $\mathbf{x}^g$, in order to make the approximation[3],

$$p(\mathbf{x}^g, \mathbf{x}^p|\omega) \rightsquigarrow p(\mathbf{x}^g - \mathbf{x}^p|\omega) \qquad (4)$$

from the offline examples present in the world set. These likelihoods are attempting to model the whole face, for both the client ($\omega = \mathcal{C}$) and imposter classes ($\omega = \mathcal{I}$). Due to the large dimensionality of the whole face $\mathbf{x}$, subspace approaches based on PCA must be employed to approximate these likelihoods from a finite world set. Additionally, it has been reported that techniques centered around linear discriminant analysis (LDA), like those seen in the Fisherface [3] algorithm, can obtain similar performance to Moghaddam and Pentland's approach. LDA based approaches employ a similar paradigm to the approach of Moghaddam and Pentland, in terms of differential appearance, although they are not framed within a strict probabilistic framework. Approaches centered around variants of LDA, have recently reported good performance on the problem of pose mismatched face recognition [4].

For the experiments in this paper, we will be assuming the client and imposter classes of the differential appear-

---

[3]It should be emphasized that we are attempting to approximate the output of the likelihood function for the purposes of classification, not the generative distribution itself. To make this difference clear, we use the $\rightsquigarrow$ to denote our approximation.
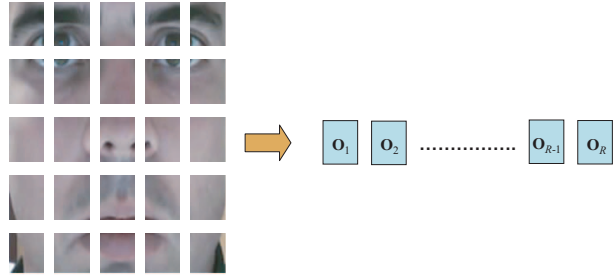

Figure 3. In this paper we will be using a patch-based representation of the face such that $\mathbf{x} = [\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_{R-1}, \mathbf{o}_R]$.

ance likelihood in Equation 4 are modeled through a normal distribution. These distributions are estimated within a subspace, found using PCA, that preserves all major modes of extra-personal variation. Constraining the distribution to lie within this subspace ensures that the covariance matrix, describing the client and imposter classes, is not rank deficient. A match-score is then obtained through the application of Bayes rule found in Equation 2.

3.3. A3: Patch Independence

Recently, Kanade and Yamada [5] proposed an effective assumption for the task of face recognition in the presence of pose mismatch. This approach is centered around the decomposition of a face image into an ensemble of sub-image patches $\mathbf{x} = \{\mathbf{o}\}_{r=1}^R$. An example of this decomposition can be seen in Figure 3.

This decomposition was motivated by hypothesized deficiencies in holistic appearance-based template matching. In holistic template matching, if we use the whole face region for comparison, it is not easy to take into account changes in appearance due to pose differences, because the appearance in a different part of a face changes in a different manner due to its complicated three-dimensional shape (e.g. the nose). By treating the face as an ensemble of independent patches we can, to some extent, circumvent this problem by learning how the discrimination of each local region of the face varies as a function of pose. Kanade and Yamada [5] proposed the following approximation,

$$p(\mathbf{x}^g, \mathbf{x}^p|\omega) \rightsquigarrow \prod_{i=1}^R p(e_i|\boldsymbol{\lambda}_{\omega_i}) \qquad (5)$$

where $e_i = \|\mathbf{o}^g_i - \mathbf{o}^p_i\|^2$ is the "sum of the squared differences" (SSD) in pixels between the gallery and probe patches at position i. The SSD measure is computed after image normalization for each patch, so it is effectively equivalent to normalized correlation. The parametric form of $\boldsymbol{\lambda} = \{\boldsymbol{\mu}, \boldsymbol{\sigma}^2\}$ is Gaussian where $\boldsymbol{\mu}$ is the mean and $\boldsymbol{\sigma}^2$ is the variance, which can be explicitly estimated offline. This

estimation is performed on both the client $\mathcal{C}$ and imposter $\mathcal{I}$ classes.

4. Face Database and Evaluation

4.1. Database

Experiments were performed on a subset of the FERET database [9], 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°, $+60^o$, $+40^o$, $+25^o$, $+15^o$, -15°, -25°, -40°, -60° respectively. In all experiments, gallery images stem from the frontal pose *ba* with probe images stemming from all other view-points. The database contains 200 subjects which were randomly divided into sets $g1$ and $g2$ both containing 90 subjects. The remaining 20 subjects were used as an imposter set for our verification experiments. The world set is used to obtain any non-client data-dependent aspects of the verification system. The evaluation and imposter sets are where the performance rates for the verification system are obtained. The $g1$ and $g2$ sets were used as the world and evaluation sets respectively. All images were geometrically normalized according to their eye-coordinates to give a cropped face image of 74×64 pixels. Examples of the gallery and probe images used in our experiments can be seen in Figure 2.

4.2. Evaluation

Verification is performed by accepting a claimant when his/her match-score is greater than or equal to Th and rejecting him/her when the match-score is less than Th, where Th is a given threshold. Verification performance is evaluated using two measures; being false rejection rate (FRR), where a true client is rejected against their own claim, and false acceptance rate (FAR), where an impostor is accepted as the falsely claimed client. The FAR and FRR measures increase or decrease in contrast to each other based on the threshold Th. The overall verification performance of a system is typically visualized in terms of a receiver operating characteristic (ROC) or detection error tradeoff (DET) curve. A simple measure for overall performance of a verification system is found by determining the equal error rate (EER) for the system, where FAR = FRR.

5. Improving Patch Independence

Due to the recency of the patch independence work [5] made with approach **A3**, we have extended the algorithm first suggested by Kanade and Yamada. We have employed this new extension so as to ensure a fair comparison between algorithms, due to the maturity and body of work in existence for approach **A1** (i.e. joint appearance) and **A2** (i.e. differential appearance).
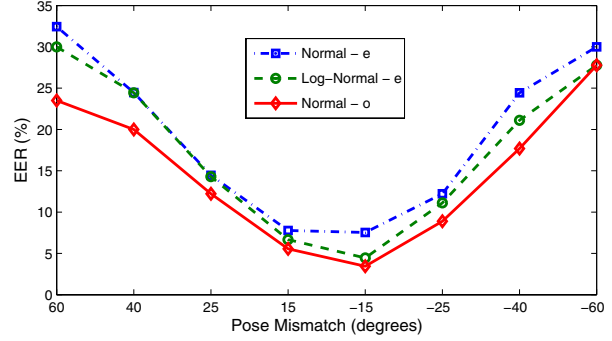

Figure 4. Results depicting the benefit of fitting a log-normal parametric form to the client and imposter's SSD values when there are small amounts of pose mismatch (i.e. $< 25^o$). There seems to be a clear advantage in modeling differential patch appearances directly, rather than relying on indirect SSD values. The patch-size for all these experiments was 8×8.

Our extension to Kanade and Yamada's algorithm was motivated by the experiments conducted in Figure 4. In these experiments, we first wanted to investigate whether there was any benefit in modeling the differential patch appearance directly; rather than relying on the more coarse SSD values. This approach attempts to make the following approximation,

$$p(\mathbf{x}^g, \mathbf{x}^p | \omega) \rightsquigarrow \prod_{i=1}^{R} p(\mathbf{o}^g_i - \mathbf{o}^p_i | \boldsymbol{\lambda}_{\omega_i}) \qquad (6)$$

Note in Equation 6 we are now modeling differences in patch appearance directly, as opposed to the indirect approach taken in Equation 5 of modeling just the SSD values. The parametric form of $\boldsymbol{\lambda}$ is assumed to be a multidimensional normal distribution. A 2D discrete cosine transform was used to preserve the 32 most energy preserving dimensions in each patch. This dimensionality reduction was performed so as to ensure the covariance matrices are well ranked.

A second aspect of Kanade and Yamada's algorithm we wanted to explore was the assumed parametric form (i.e. a normal distribution) of the client and imposter SSD values. In row (a) of Figure 5, one can see a set of histograms for the offline client and imposter distributions of the SSD measure e for both the gallery and probe patch locations i across a number of pose mismatches. From these small set of visual examples in Figure 5 it is clear that the distributions are not adequately described parametrically by a normal distribution; as can be seen by the client and imposter approximations in rows (b) and (c) respectively. It would make sense to perhaps use the raw histograms obtained from the offline world set. However in practice, using the raw histograms produced poor performance, during verification, due to their noisy nature.
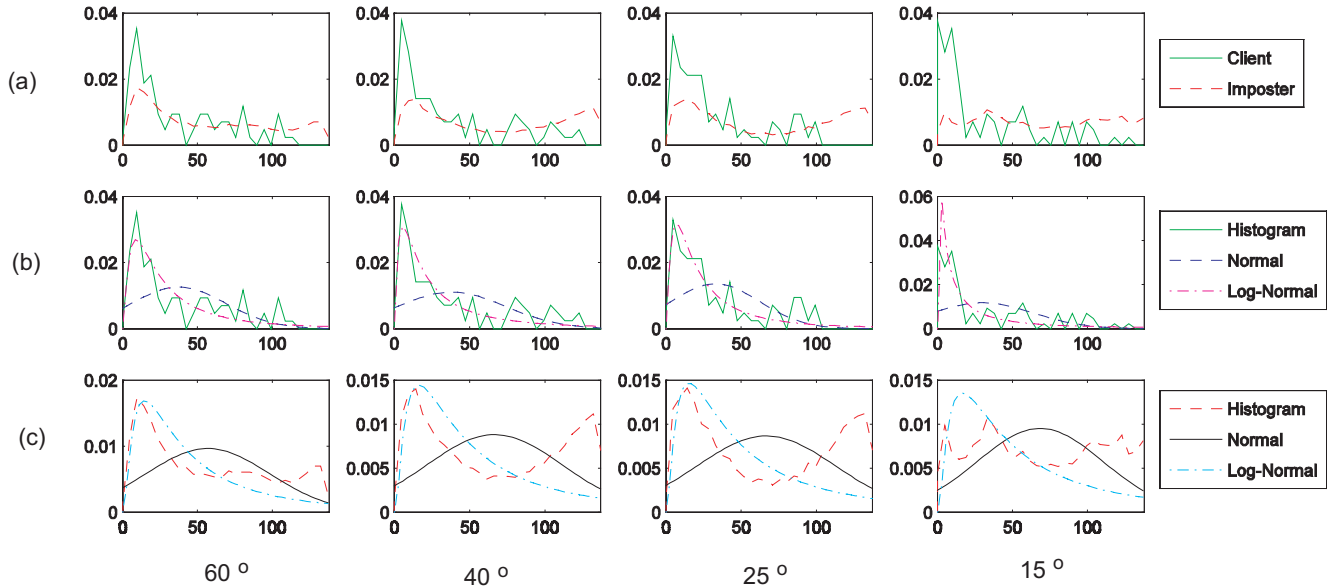
Figure 5. For all plots in this figure the x-axis denotes the SSD measure e and the y-axis denotes the density approximation $p(e|\omega)$ for the given class $\omega = \{\mathcal{C}, \mathcal{I}\}$ being modeled. Row (a) depicts the histograms for the the client and imposter classes for an individual patch i across four different pose mismatches (i.e. $60^o, 40^o, 25^o, 15^o$). The individual patches i for each pose were chosen based on the client and imposter histograms that had the largest divergence. Rows (b) and (c) denote the parametric estimations of the client and imposter distributions respectively. In our experiment we attempted to fit a normal and log-normal distribution. One should note that employing a normal distribution for either the client or imposter distributions resulted in a poor fit.

To circumvent the problems of poor fit, with respect to the normal distribution, and of noise, with respect to the histogram fit, we propose the employment of the log-normal distribution,

$$p(e|\boldsymbol{\lambda}) = \frac{1}{\gamma\sqrt{2\pi e}} \exp(-\frac{\|\log(e) - \nu\|^2}{2\gamma^2}) \qquad (7)$$

The parametric values $\boldsymbol{\lambda} = \{\nu, \gamma^2\}$ are found in a similar explicit manner to the mean and variance values of the normal distribution; with the $\log$ operator being used before one estimates the log-mean $\nu = E\{\log(e)\}$ and log-variance $\gamma^2 = E\{log(e)^2\} - E\{log(e)\}^2$ parameters.

The results presented in Figure 4 demonstrates that it is advantageous to employ a log-normal, rather than normal, distribution for modeling patch SSD values. This result is especially noticeable where there are smaller pose mismatches. A possible reason for this result can be attributed to the increased likelihood that the client and imposter distributions will be skewed towards a zero SSD value when there is less pose mismatch. The most important result in Figure 4 is the clear advantage seen in modeling the differential patch observations directly, rather than relying on the indirect SSD values. This extension to the algorithm representing approach **A3** is employed throughout the rest of our experimental results.

6. Assessing Approaches

In Figure 6 one can see a performance breakdown of algorithms representing the three approaches **A1-A3**, referring to modeling the *whole joint appearance*, *whole differential appearance*, and *patch differential appearance* respectively gallery-probe face image pairs. We also compared performance against the Eigenface [2] algorithm, as this is a common benchmark used in literature. Algorithm **A1** seems to obtain the best performance overall, in comparison to algorithms **A2** and **A3**; with all three algorithms outperforming the baseline Eigenface algorithm.

In Figure 7, we conducted an additional experiment where we tested the performance of algorithms **A1** and **A2** for the situation where the gallery image is "badly" misaligned with the probe image. We synthetically created this misalignment by performing a 180° circular shift on the gallery image in the x and y directions. An example of this synthetic misalignment can be seen at the bottom of Figure 7. It must be emphasized, for these experiments, that the circular shift operation was applied to both the offline and online gallery images; requiring the likelihood functions for algorithms **A1** and **A2** to be re-estimated. Interestingly, there was no noticeable degradation in performance for algorithm **A1**, whereas algorithm **A2** suffered catastrophic degradation in comparison to the original results seen in Figure 6.
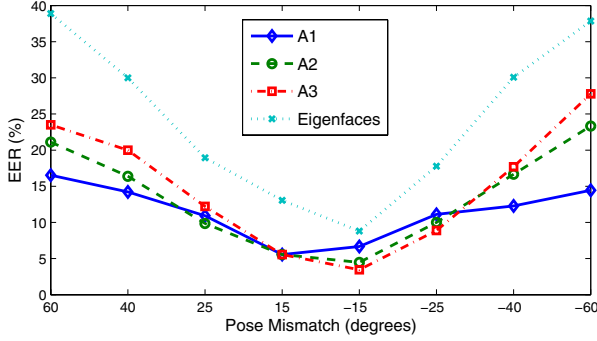
Figure 6. Results demonstrating the breakdown in performance for the three approaches **A1-A3**, referring to modeling the *whole joint appearance*, *whole differential appearance*, and *patch differential appearance* respectively. Results demonstrate that there is benefit in modeling the whole joint appearance, especially in the presence of a large pose mismatch (i.e. greater than 40°).
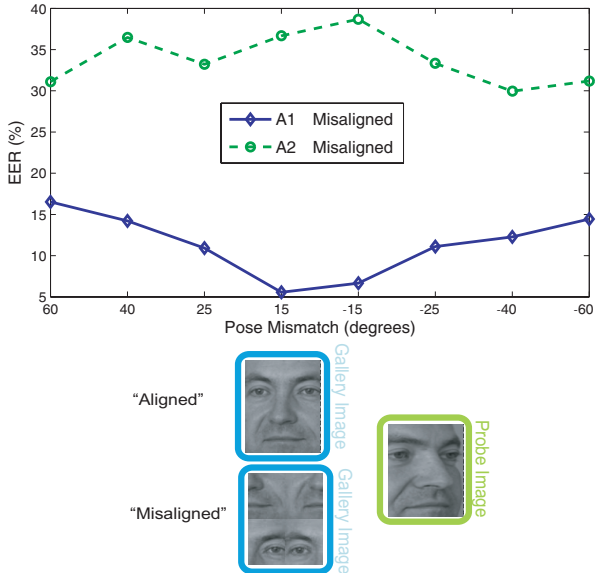

Figure 7. Demonstration of how algorithms that model joint appearance (**A1**) rather than differential appearance (**A2**) are less prone to the effects of "bad" alignment between gallery and probe images. All results in this figure were carried out on misaligned gallery images. Refer to Figure 6 for the aligned performance of algorithms **A1** and **A2**. Note that there is minimal difference in performance between the aligned and misaligned experiments for **A1**. However, there is a catastrophic drop in performance for algorithm **A2** for the misaligned experiment.

An immediate observation one can make about the experimental results in Figures 6 and 7, is that any algorithm that relies on modeling differential appearance, whether at the holistic or patch level, intrinsically relies on "some level" of alignment between the gallery and probe images. As pose mismatch increases, the alignment of gallery and

probe images tends to degrade; resulting in poorer verification performance. Assuming independence between patches, for differential appearance, causes further degradation, albeit slight, in performance as pose mismatch increases.

7. Learning Patch Dependence

It seems obvious from the experiments conducted in previous sections, that alignment is a problem when one is employing a model based on differential appearance in the presence of large pose mismatch; whether it be at a holistic or patch level. However, techniques that attempt to model the joint appearance, of a gallery and probe image, employ large assumptions to make the estimation of such a likelihood function feasible. Algorithms, like the one employed in **A1** (e.g. Eigen Light-Fields, TensorFaces), assume the joint appearance of a set of gallery and probe images is governed by a single Gaussian distribution.

In this section we propose an approach, based on a patch representation of the face, that does not rely on such a constraining assumption. In our approach we make the following approximation,

$$p(\mathbf{x}^g, \mathbf{x}^p | \omega) \leadsto \prod_{i=1}^{R} p(\mathbf{o}^g{}_i, \mathbf{x}^p | \boldsymbol{\lambda}_{\omega_i}) \qquad (8)$$

where $\mathbf{o}^g{}_i$ refers to an image patch, at position i, within the gallery image, and $\mathbf{x}^p$ refers to the whole appearance of the probe image. In this approach we attempt to learn the gallery patch dependence with respect to the whole probe image appearance. Unlike the **A3** algorithm, employed in previous sections, this approximation does not make any assumptions about the alignment of gallery and probe images. However unlike the **A1** algorithm, our approach takes advantage of a patch-decomposition of the gallery image. Each gallery-patch/probe-whole appearance likelihood function is assumed to be Gaussian. PCA is employed to reduce the total dimensionality of the gallery-patch/probe-whole representation. Each gallery-patch and probe-whole representation is normalized for energy, so as limit a bias towards modeling the probe-whole appearance solely. In our approach, unlike **A1**, we are also able to approximate an imposter likelihood function due to the diminished dimensionality of the representation. A match-score is obtained through the application of Bayes rule in Equation 2.

The performance of our method can be seen in Figure 8 in comparison to the **A1** algorithm which showed the best overall performance in Figure 6. One can see that our approach demonstrates considerable improvement over the **A1** algorithm, especially in the presence of large pose mismatches. A detection error tradeoff (DET) curve can be seen in Figure 9 for the extreme case of 60° angle mismatch between the probe and frontal gallery image. For
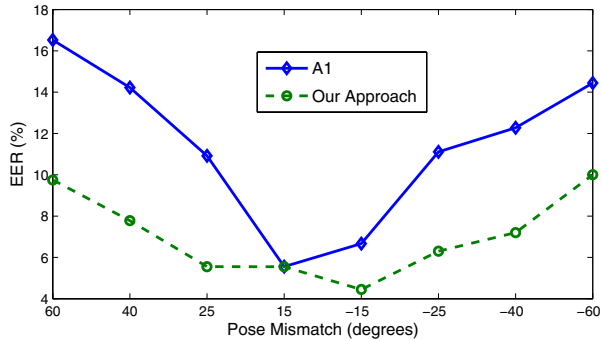
Figure 8. Results demonstrating the improved verification performance, in comparison to the **A1** algorithm, for our proposed method. Note that our algorithm shows considerable improvement over the **A1** algorithm for almost all pose mismatches.
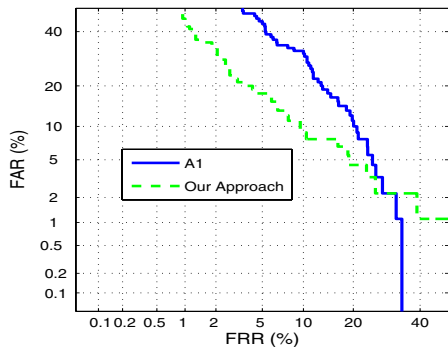

Figure 9. A detection error tradeoff curve (DET) is depicted here for the extreme case of 60° mismatch between the probe and frontal gallery images. One can see that our method outperforms the **A1** algorithm across nearly every operating point, in terms of FAR and FRR.

nearly every operating point, our method outperforms the **A1** algorithm in terms of FAR and FRR.

8. Conclusions

In this paper we have conducted experiments associated with three major approaches in pose mismatched face recognition literature. This comparative analysis demonstrated that approaches **A2** and **A3**, that rely on differential appearance at the holistic and patch level respectively, perform poorly in the presence of large rotation (i.e. $> 40^o$) about the vertical axis. Through additional experimentation we demonstrated that this poor performance can largely be attributed to the poor alignment of the gallery and probe images as pose mismatch increases. As a result, algorithms that do not make assumptions about the alignment of the gallery-probe image pairs, such as approach **A1** which modeled the joint appearance, tended to obtain improved performance.

An auxiliary benefit of our work also resulted in an extension to the original **A3** approach, first proposed by Kanade and Yamada [5]. This novel extension demonstrated improved performance by employing a log-normal, rather than, normal distribution for modeling SSD values. Further, we demonstrated best performance by modeling the differential patch appearances directly.

Finally, we proposed an algorithm that unifies the benefits of approaches **A1** and **A3** by learning the dependencies of gallery patches with the whole probe image appearance. This approach has two advantages. First, it makes no assumption about the alignment between the gallery and probe image pairs; allowing it to deal with large pose mismatch. Secondly, it allows for a richer modeling of the joint appearance by decomposing the gallery image into an ensemble of statistically independent patches. Our approach out-performed all other approaches tested in our experiments. The performance of our algorithm in large pose mismatch was especially encouraging.

References

[1] V. Blanz and T. Vetter, "Face Recognition Based on Fitting a 3D Morphable Model," *IEEE Trans. PAMI*, vol. 25, no. 9, 2003.

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

[3] 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.

[4] T. Kim and J. Kittler, "Locally Linear Discriminant Analysis for Multimodally Distributed Classes for Face Recognition with a Single Model Image," *IEEE Trans. PAMI*, vol. 27, pp. 318–327, March 2005.

[5] T. Kanade and A. Yamada, "Multi-subregion Based Probabilistic Approach Towards Pose-Invariant Face Recognition," in *IEEE International Symposium on Computational Intelligence in Robotics Automation*, vol. 2, pp. 954–959, 2003.

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

[7] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object recognition," *IEEE Trans. PAMI*, vol. 19, no. 7, pp. 696–710, 1997.

[8] 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.

[9] 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.

[10] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Trans. PAMI*, vol. 15, no. 10, pp. 1042–1052, 1993.

[11] A. M. Martínez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Trans. PAMI*, vol. 24, no. 6, pp. 748–763, 2002.