Using Overlapping Distributions to Deal with Face Pose Mismatch

Simon Lucey Tsuhan Chen

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

Abstract

Representing the face as a distribution of freely moving patches, which we refer to as a "free-parts" representation, has recently demonstrated some benefit in the task of face verification. This benefit can be largely attributed to the representation's natural ability to deal with local appearance variation within the face. Hitherto, a major limitation that has hindered the wider adoption of this type of facial representation, for the task of face verification, has been its poor ability to take advantage of prior knowledge concerning mismatches in context; such as pose. This paper goes some way to alleviating these limitations by making two novel contributions: (i) Demonstrating that free-parts distributions of a client's face for different poses overlap to such a degree that a considerable amount of discrimination is preserved in the intersection. (ii) Through the off-line estimation of subject-independent pose dependent priors, an alternative to the canonical log-likelihood measure can be employed that takes advantage of this intersection and is less sensitive to mismatch in the presence of pose variation.

1 Introduction

Representing a face as an ensemble of local image patches, rather than a holistic monolith of pixels, has some inherent advantages when trying to match faces in real-world scenarios. This representation can be especially beneficial in the presence of heterogeneous mismatches (i.e. mismatches that do not occur equally across all parts of the image) such as those occurring from differing pose, shadow, occlusion, expression, etc. In the presence of these types of heterogenous mismatches any benefit gained from modeling the global dependencies in a face image is usually lost by the overpowering effect of the mismatch in appearance globally. However, by trying to match an image based purely on its local appearance (i.e. decomposing a single face image into many sub-face image patches) one can alleviate this effect, as the effect of the mismatch is usually not occurring equally across the image (i.e. heterogeneous). A specific representation that embraces this concept is a "free-parts" representation as the patches within the face are able to "freely" move to varying extents by relaxing the position/structure of patches within the face image. A high-level graphical example of how one can obtain a free-parts distribution from a client image is depicted in Figure 1. This representation has already met with some

success in frontal-face recognition [3], in the presence of alignment and expression mismatch.

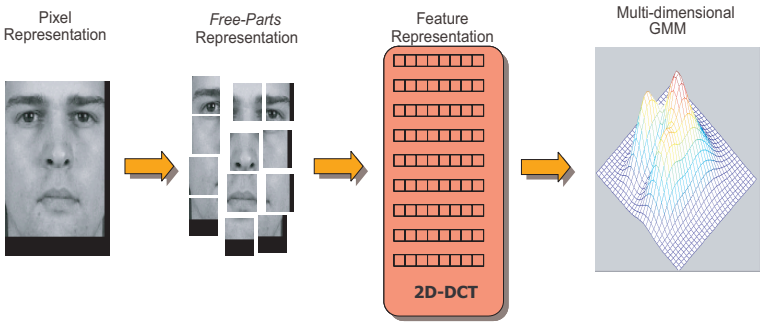


Figure 1: A graphical depiction of the process involved in obtaining a "free-parts" Gaussian mixture model (GMM) distribution from a client image. Note: in a free-parts representation positional information of where patches in the image are located are ignored, allowing the patches to "freely" move.

Hitherto, a major limitation that has hindered the wider adoption of this type of facial representation, for the task of face verification, has been its poor ability to take advantage of prior knowledge concerning mismatches in context; such as pose. In this paper we shall refer to this prior knowledge in the form of a world class $\mathcal{W}_i$, where i is an index to a pre-defined number of discrete poses. The world class is representative of the general population of faces for the pose i, whose distribution model is usually estimated from a large ensemble of non-client faces that are separate to those used in the face verification task. In our work we assume we have world models for all possible poses that will be encountered in evaluation, which is a reasonable assumption as the world models are independent of the subjects being verified and can be estimated offline. A fundamental question being asked in this paper is: how can one incorporate the prior knowledge of world classes $\mathcal{W}_i, \forall i$ given there is a mismatch in pose between a client's gallery and the claimant's probe image?

In this paper we make the highly novel claim that prior knowledge of the classes $\mathcal{W}_i, \forall i$ can be successfully leveraged to take advantage of the heterogenous nature of pose change on a face's local appearance (i.e. some patches of the face will be more prone to 3D appearance variation, such as the nose, etc., than other patches). Specifically, we propose one can categorize patches in two images of the same client for differing viewpoints i and j in three ways: (i) a patch is common to view-points i and j, (ii) a patch is common only to view-point i, and (iii) a patch is common only to view-point j. A depiction of this categorization can be seen in Figure 2. Work by and Kanade and Yamada [1] employed a similar categorization of patches in their work for pose mismatched face recognition, although their work did not deal with distributions of the patches themselves.

Through the framework of this categorization we present two novel contributions. First, that a large amount of discriminative information exists, and can be empirically measured, in the overlapping portions (i.e. the intersection class) of a free-parts distribution for poses i and j of the same client. Second, that much of this discriminative information can be extracted, using our prior knowledge of the world models, through the

emphasis and de-emphasis of intersecting and non-intersecting patches respectively. We presents our results on the pose portion of the FERET [4] database.

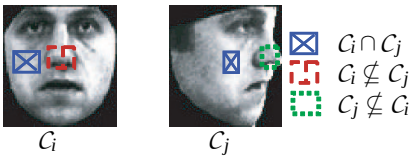


Figure 2: A depiction of the three ways to categorize a patch, given images from the view-points i and j for the same subject class $\mathcal{C}$. Specifically, one can categorize a patch as: (i) common to i and j ($\mathcal{C}_i \cap \mathcal{C}_j$), (ii) common to only i ($\mathcal{C}_i \nsubseteq \mathcal{C}_j$), and (iii) common to only j ($\mathcal{C}_j \nsubseteq \mathcal{C}_i$). For all experiments in this paper the gallery pose i will be frontal, with all probe poses j being non-frontal. Note: patches common to i and j will have different aspect ratios at each view-point.

2 Problem Definition

Borrowing upon set theory we can formalize the categorizations of patches in Section 1 as the classes: (i) $\mathcal{C}_i \cap \mathcal{C}_j$ (i.e. the intersection), (ii) $\mathcal{C}_i \nsubseteq \mathcal{C}_j$, and (iii) $\mathcal{C}_j \nsubseteq \mathcal{C}_i$ (i.e. the non-intersections). A graphical interpretation of these classes in terms of conditional pdfs can be seen in Figure 3.

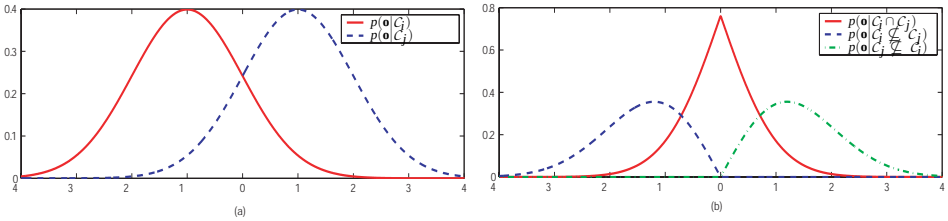


Figure 3: A graphical example of how the pdfs for $\mathcal{C}_i$ and $\mathcal{C}_j$ (depicted as Gaussians in (a) in this example) can be decomposed into (i) $\mathcal{C}_i \cap \mathcal{C}_j$, (ii) $\mathcal{C}_i \nsubseteq \mathcal{C}_j$ and (iii) $\mathcal{C}_j \nsubseteq \mathcal{C}_i$ as seen in (b). Where: (i) represents the class of observations that intersect with $\mathcal{C}_i$ and $\mathcal{C}_j$, (ii) represents the class of observations of class $\mathcal{C}_i$ that do *not* intersect with $\mathcal{C}_j$, and (iii) represents the the class of observations of class $\mathcal{C}_j$ that do *not* intersect with $\mathcal{C}_i$.

In our work we assume that we have parametric representations of $\mathcal{C}_i$ and $\mathcal{W}_i$, for pose i, that encapsulates the sufficient statistics (i.e. $\lambda_{\mathcal{C}_i}, \lambda_{\mathcal{W}_i}$) of each class. For the empirical portion of this paper we will make approximations to these parametric forms using Gaussian mixture models (GMMs)[1]; for more information on how we obtain the free-parts representation of an image and then estimate GMMs please refer to appendix Sections A and B respectively. In practice one typically has a client model for only a *single* viewpoint, which is usually mismatched to the viewpoint of the claimant images. However,

[1]For our experimental work we used GMMs with 32 mixture components and diagonal covariance matrices. The dimensionality of the 16×16 observations patches, after a two-dimensional discrete cosine transform (DCT) based feature extraction, is $d = 35$.

we do have access to world models for all viewpoints that will be seen in the claimant images.

We shall denote patches from the claimant image for pose j as $\mathbf{D}_j = \{\mathbf{o}_r\}_{r=1}^{R}$ where $\mathbf{o}$ refers to an individual observation image patch and R is the number of patches in the face image. One can obtain the client specific posterior probability for a single observation as,

$$P(\mathcal{C}_i|o) = \frac{p(\mathbf{o}|\mathcal{C}_i)}{p(\mathbf{o}|\mathcal{C}_i) + p(\mathbf{o}|\mathcal{W}_i')} \tag{1}$$

where we assume equal priors between client and world classes. The task of face verification, for a free-parts representation, can now be defined as,

$$\mathcal{L}(\mathbf{D}_j|\mathcal{C}_i) \underset{\text{accept}}{\overset{\text{reject}}{\lessgtr}} Th \tag{2}$$

where,

$$\begin{aligned} \mathcal{L}(\mathbf{D}_j|\mathcal{C}_i) &= E\{\log P(\mathcal{C}_i|\mathbf{o})|\mathcal{W}_j\} \\ &= \frac{1}{R}\sum_{r=1}^{R} \log P(\mathcal{C}_i|\mathbf{o}_r) \end{aligned} \tag{3}$$

One might note that the notation employed in Equation 3 departs from the canonical method of expressing the log-likelihood for a set of independent observations. In this formulation we weight each observation in $\mathbf{D}_j$ equally when calculating the expectation, so we say the expectation has been taken with respect to $\mathcal{W}_j$; as all observations (i.e. patches) in $\mathbf{D}_j$ are a subset of the world class $\mathcal{W}_j$. When a patch in the claimant image is *not* a subset of the world class of which the expectation is being taken with respect to, for example if we were taking the expectation with respect to $\mathcal{W}_i' \cap \mathcal{W}_j$, a re-weighting scheme must be employed as we shall explore in Section 5. The threshold Th is specified for a desired false accept rate (FAR) or false reject rate (FRR).

The log-likelihood measure in Equation 3, which we shall refer to as the geometric mean (GM) log-likelihood, is optimal if $i = j$ and we assume the observations in $\mathbf{D}_j$ are identically and independently distributed (i.i.d.). For the task of face recognition it is common however, to encounter $i \neq j$ (i.e. the claimant's pose is not similar to the client model's pose) resulting in a mismatch, making the log-likelihood in Equation 3 sub-optimal as will be discussed further in Section 5. As one never has access to $\mathcal{C}_j$ if $i \neq j$ the task for the rest of this paper is to estimate a useful log-likelihood measure based solely on our knowledge of the classes $\mathcal{C}_i, \mathcal{W}_i'$ and $\mathcal{W}_j$.

3 Face Database and Experiments

Experiments were performed on a subset of the FERET database [4], 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. 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 interchangeably as the world and evaluation sets.

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 and angle 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 due to a stretching of the image in the y-axis. In our work we chose to employ the distance from the eye line to the nose tip vertically to remedy the stretching problem. The final geometrically normalized cropped faces formed a 98×115 array of pixels. The face verification task is the binary process of accepting or rejecting the identity claim (i.e. the log-likelihood measure) made by a subject under test. 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.

4 Modeling the Intersection Class

One of the major advantages of employing a free-parts representation over more holistic representations (e.g. Eigenfaces [6]) of the face is the additional avenues it provides for dealing with mismatch. By modeling the face as a cloud of image patch observations (i.e. a distribution), one can take advantage of the heterogeneous nature of pose mismatches at a patch level (i.e. some patches within a face from pose i will be of more use than others when trying to match with a face from a differing pose j). From a probabilistic view point this translates to modeling the intersection class between two poses. Such an intersection would not be possible with holistic type representations that represent the face as a single observation, as the global appearance will always differ (i.e. not intersect) between viewpoints. An example of this concept of intersection can be seen in Figure 3 where one can see that a client's distribution for two poses can be divided into,

$$p(\mathbf{o}|\mathcal{C}_i) = P(\mathcal{C}_i \cap \mathcal{C}_j)p(\mathbf{o}|\mathcal{C}_i \cap \mathcal{C}_j) + [1 - P(\mathcal{C}_i \cap \mathcal{C}_j)]p(\mathbf{o}|\mathcal{C}_i \nsubseteq \mathcal{C}_j) \tag{4}$$

where $p(\mathbf{o}|\mathcal{C}_i \cap \mathcal{C}_j)$ is the pdf for observations common to $\mathcal{C}_i$ and $\mathcal{C}_j$. Similarly, $p(\mathbf{o}|\mathcal{C}_i \nsubseteq \mathcal{C}_j)$ is the pdf for observations *not* intersecting with $\mathcal{C}_j$ given that they stem from $\mathcal{C}_i$. Additionally, $P(\mathcal{C}_i \cap \mathcal{C}_j)$ is a prior employed to ensure that $p(\mathbf{o}|\mathcal{C}_i \cap \mathcal{C}_j)$ integrates to unity.

In our work we make the simplifying assumption that all observations (i.e. image patches) stemming from each class are i.i.d., as this assumption has good generalization properties when having to do face recognition in the presence of large pixel appearance variation [3]. In this paper we make the additional simplification that this i.i.d. assumption extends across poses (i.e. there is no conditional dependence between observations of the same subject between different poses)[2] so as to concentrate solely on leveraging prior knowledge of the overlap between contexts for improved face verification performance.

[2] Although in future work we would like to incorporate conditional dependence between mismatches into our framework for improved performance.

4.1 How much overlap?

A suitable question to ask before embarking upon the rest of this paper is: how much discriminating information is contained in the intersection (if there is any intersection?) of free-parts distributions between poses? In this paper we have formulated a novel technique that can obtain an empirical answer to this question. To answer this question we make the assumption that we have access to models for the classes $\mathcal{C}_i$ *and* $\mathcal{C}_j$. This is *not* a viable assumption in practice, as one only ever has access to $\mathcal{C}_i$ but not $\mathcal{C}_j$; where *i* and *j* are the gallery and probe poses respectively. We are entertaining this assumption however, for this section to gain a measure of how much discriminative information exists in the class $\mathcal{C}_i \cap \mathcal{C}_j$.

Given that we have access to estimates of the parametric models for classes $\mathcal{C}_i$ and $\mathcal{C}_j$ one can define the intersection class pdf as,

$$P(\mathcal{C}_i \cap \mathcal{C}_j)p(\mathbf{o}|\mathcal{C}_i \cap \mathcal{C}_j) = \begin{cases} P(\mathcal{C}_i)p(\mathbf{o}|\mathcal{C}_i), & \text{If } \eta_{j/i}(\mathbf{o}) > 1 \\ P(\mathcal{C}_j)p(\mathbf{o}|\mathcal{C}_j), & \text{otherwise} \end{cases} \tag{5}$$

where,

$$\eta_{j/i}(\mathbf{o}) = \frac{P(\mathcal{C}_j)p(\mathbf{o}|\mathcal{C}_j)}{P(\mathcal{C}_i)p(\mathbf{o}|\mathcal{C}_i)} \tag{6}$$

the intersecting world pdf $p(\mathbf{o}|\mathcal{W}_i \cap \mathcal{W}_j)$ can be estimated in a similar manner. Substituting $p(\mathbf{o}|\mathcal{C}_i \cap \mathcal{C}_j)$ and $p(\mathbf{o}|\mathcal{W}_i \cap \mathcal{W}_j)$ into Equation 1 to obtain $P(\mathcal{C}_i \cap \mathcal{C}_j|\mathbf{o})$; we can obtain the following log-likelihood measure,

$$\mathcal{L}(\mathbf{D}_j|\mathcal{C}_i \cap \mathcal{C}_j) = E\{\log P(\mathcal{C}_i \cap \mathcal{C}_j|\mathbf{o})|\mathcal{W}_j\} \tag{7}$$

This log-likelihood measure, which we shall refer to as the intersecting geometric mean (IGM) log-likelihood measure, is optimal from the perspective that it gives us a measure of performance for claimant observations in $\mathbf{D}_j$ that intersect with both $\mathcal{C}_i$ and $\mathcal{C}_j$. We can interpret the IGM results as a lower bound on verification performance, when trying to take advantage of the intersection of mismatched distributions from the same client. One can see a comparison between the baseline GM and the lower bound IGM log-likelihood measures in Figure 4.

Figure 4 demonstrates two important facts concerning free-parts face verification in the presence of pose mismatch. First, that there is a large number of common (i.e. intersecting) observations between a client's $\mathcal{C}_i$ and $\mathcal{C}_j$ models. Second, that the non-intersecting observations drastically affect performance on the larger mismatched poses as can be seen from the poor baseline performance of the GM measure. The rest of this paper will be dedicated to measures that can try and approximate the performance of the IGM log-likelihood measure given we do *not* have access to $\mathcal{C}_j$.

5 Robust Measures

In previous work [2] concerning the effective combination of conditionally independent classifiers it has been shown that the multiplication of conditionally independent probabilities is extremely sensitive to mismatch errors due to the geometric mean property (i.e. the smallest values in a numeric sequence will dominate the geometric mean). In this paper we propose a similar effect occurs when attempting to calculate the baseline

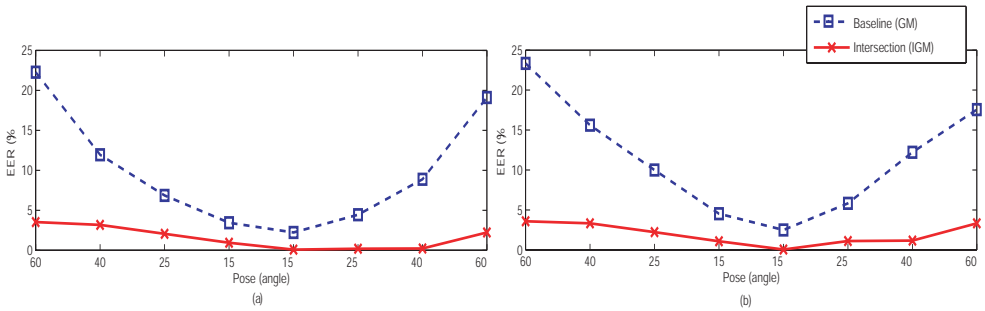


Figure 4: A comparison between the baseline GM (3) and theoretical IGM (7) log-likelihood measures, demonstrating that considerable discriminating information exists in the intersection of client distributions for mismatched pose. Note: (a) depicts results for evaluations on the *ga* set and (b) depicts results for the *gb* set.

GM log-likelihood measure, where probabilities for non-intersecting observations in the presence of a heterogeneous mismatch dominate because of their small values.

To alleviate this problem we propose a number of log-likelihood measures that are far less sensitive to small probabilities stemming from mismatch. Our first proposed measure is the arithmetic mean (AM) log-likelihood measure,

$$\begin{aligned} \mathcal{L}(\mathbf{D}_j|\mathcal{C}_i) &= \log E\{P(\mathcal{C}_i|\mathbf{o})|\mathcal{W}_j\} \\ &= \log \frac{1}{R}\sum_{r=1}^{R} P(\mathcal{C}_i|\mathbf{o}_r) \end{aligned} \tag{8}$$

This measure is motivated by work from Kittler et al. [2] who demonstrated how an arithmetic sum of conditionally independent probabilities containing error is far less sensitive than a geometric sum, which is considered optimal if those same probabilities do not contain errors. Additionally, we can incorporate our prior knowledge of the intersection of $\mathcal{W}_i$ and $\mathcal{W}_j$ into the measure,

$$\begin{aligned} \mathcal{L}(\mathbf{D}_j|\mathcal{C}_i) &= \log E\{P(\mathcal{C}_i|\mathbf{o})|\mathcal{W}_i \cap \mathcal{W}_j\} \\ &= \log \sum_{r=1}^{R} \frac{\delta(\mathbf{o})P(\mathcal{C}_i|\mathbf{o}_r)}{\delta(\mathbf{o})} \end{aligned} \tag{9}$$

where,

$$\delta(\mathbf{o}) = \frac{p(\mathbf{o}|\mathcal{W}_i \cap \mathcal{W}_j)}{p(\mathbf{o}|\mathcal{W}_j)} \tag{10}$$

Note, the expectation in Equation 9 is taken with respect to $\mathcal{W}_i \cap \mathcal{W}_j$, which requires an unequal weighting of observations in $\mathbf{D}_j$. The weighting factor $\delta(\mathbf{o})$ described in Equation 10 is to leverage our knowledge that observations existing in $\mathcal{W}_i \cap \mathcal{W}_j$ will be less susceptible to mismatch than those observations not lying in the intersection. We refer to the measure described in Equations 9 and 10 as our world intersection arithmetic mean (WIAM) log-likelihood measure.

Our third proposed measure endeavors to desensitize the effect of mismatch on our scores in an alternate manner through the approximation,

$$p(\mathbf{o}|\mathcal{C}_i \cup \mathcal{C}_j) \approx P(\mathcal{C}_i)p(\mathbf{o}|\mathcal{C}_i) + [1 - P(\mathcal{C}_i)]p(\mathbf{o}|\mathcal{W}_j) \tag{11}$$

where we assume $P(\mathcal{C}_i) = 0.5$ for our current framework. This approach attempts to ensure that any observations that are unlikely in $\mathcal{C}_i$ but likely in $\mathcal{W}_j$ do not adversely affect the aggregate likelihood. Employing Equations 1 and 11 we can obtain the approximation $\hat{P}(\mathcal{C}_i \cup \mathcal{C}_j|\mathbf{o})$[3] which can be used to form the likelihood,

$$\mathcal{L}(\mathbf{D}_j|\mathcal{C}_i) = E\{\log \hat{P}(\mathcal{C}_i \cup \mathcal{C}_j|\mathbf{o})|\mathcal{W}_j\} \tag{12}$$

which we shall refer to as the world union geometrical mean (WUGM) log-likelihood measure.

6 Results and Discussion

In Figure 5 one can see verification results for our baseline GM (3) and proposed AM (8), WIAM (9), WUGM (11) log-likelihood measures. Results are presented on both the *ga* and *gb* sets of the FERET database, in Figures 5 (a) and (b) respectively. In these results one can see that all three proposed measures perform better than the baseline GM measure for nearly all poses. One can additionally see that our proposed WIAM measure, which attempts to desensitize the effect of mismatch error on the overall likelihood and emphasize observations that lie in the intersection $\mathcal{W}_i \cap \mathcal{W}_j$, has better performance than the AM and WUGM measures across all poses. This superior performance can be attributed to the AM and WUGM measures not taking advantage of the prior knowledge we have about the intersection between world models for differing poses, instead concentrating solely on desensitizing the effect of mismatch errors.

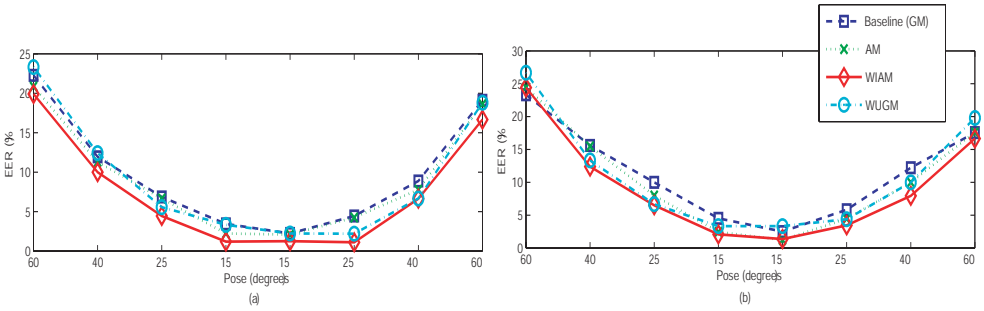


Figure 5: Depiction of verification results for the (a) *ga* and (b) *gb* sets for our baseline GM (3) and proposed AM (8), WIAM (9), WUGM (11) log-likelihood measures. One can see that our proposed WIAM measure gives improved performance in nearly all circumstances.

In conjunction to the work we have presented in this paper on robust log-likelihood measures we have also investigated alternate strategies for gaining estimates of $p(\mathbf{o}|\mathcal{C}_i \cap \mathcal{C}_j)$

[3]We use the notation $\hat{P}(\mathbf{o})$ to denote an approximation to the real probability $P(\mathbf{o})$.

directly, given that we only have prior knowledge of $\mathcal{C}_i$, $\mathcal{W}_i$ and $\mathcal{W}_j$. In this work we proposed a GMM estimation strategy based on the EM-algorithm and our prior knowledge of $\mathcal{W}_i$ and $\mathcal{W}_j$, which would weight observations in the client set $\mathcal{C}_i$ based on how likely they are in $\mathcal{W}_i \cap \mathcal{W}_j$. This approach worked well on synthetic problems where we were able to employ a $\mathcal{C}_i$ with a very large number of observations. However, the approach fell apart in the more realistic situation where one has a meager number of observations describing the client set $\mathcal{C}_i$. This poor result can be attributed to the very nature of the intersection class between two distributions, as the observations lying in the intersection are inherently of low likelihood resulting in very erroneous estimates of $p(\mathbf{o}|\mathcal{C}_i \cap \mathcal{C}_j)$ given that one only has access to $\mathcal{C}_i$.

One can see from the results in Figure 5 that there is a definite advantage in employing our alternate log-likelihood measures in the presence of pose mismatch. However, for substantial pose mismatch (i.e. $\theta \geq +/-40^o$) the improvement seen is substantially less than those performances seen for the theoretical IGM (7) measure in Section 4.1. We believe that this poor result can be explained in a similar manner to the previously described problems of trying to estimate a client's intersecting model directly. Due to the claimant observation set $\mathbf{D}_j$ being of a finite size, the proportion of observations that have a positive effect on the model $p(\mathbf{o}|\mathcal{C}_i)$ is considerably less than $p(\mathbf{o}|\mathcal{C}_i \cap \mathcal{C}_j)$. In future work we believe modeling the dependencies found in the joint distribution $p(\mathbf{o}|\mathcal{W}_i, \mathcal{W}_j)$ may overcome this limitation, as one can employ non-intersecting observations, which there are inherently more of, to aid in the estimation of $p(\mathbf{o}|\mathcal{C}_i \cap \mathcal{C}_j)$ given we only have access to $\mathcal{C}_i$.

7 Conclusions

In this paper we have presented a novel framework that attempts to leverage prior knowledge about the intersection class, for different pose mismatches, to gain improved face recognition performance. This approach is particularly novel as it attempts to emphasize and de-emphasize image patches in the face based, on prior knowledge concerning the overlap of free-parts distributions for mismatched poses. In this work we demonstrated how one can define the intersection class between two overlapping distributions, and additionally gave an empirical measure of the amount of discriminative information existing in this intersection for the task of verification. This high discrimination is made possible through the employment of a free-parts representation of the face; that encourages the viewing of a face image not as a single point, but as a cloud of patch observations best described as a distribution.

We have additionally demonstrated that there are problems in taking advantage of the discriminative information contained in this intersection due to the client class $\mathcal{C}_j$ being unseen during evaluation, that is we only have access to the client class $\mathcal{C}_i$ and the prior world classes $\mathcal{W}_i$ and $\mathcal{W}_j$ where i and j are the discrete viewpoints of the client and claimant face images respectively. To combat these problems we have proposed a novel robust log-likelihood measure that: a) de-sensitizes the effect of mismatch on individual image patches in the overall log-likelihood measure for a claimant, and (b) re-weights the contribution of patches for the overall log-likelihood measure based on whether they lie in the intersection of the world models $\mathcal{W}_i$ and $\mathcal{W}_j$. In future work we believe that additional performance improvement can be attained through the joint modeling of the world classes $\mathcal{W}_i$ and $\mathcal{W}_j$.

References

[1] 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*, pages 954–958, Kobe, Japan, July 2003.

[2] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE Trans. PAMI*, 20(3):226–239, March 1998.

[3] 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)*, volume II, pages 855–861, Washington D.C., June 2004.

[4] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. PAMI*, 10(22):1090–1104, 2000.

[5] C. Sanderson and K. Paliwal. Fast features for face authentication under illumination direction changes. *Pattern Recognition Letters*, 24(14):2409–2419, 2003.

[6] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 1991.

A Free-parts Representation

To gain a free-parts representation 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 two-dimensional discrete cosine transform (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 [3, 5].

B GMM Estimation

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

$$p(\mathbf{o}|\lambda) = \sum_{m=1}^{M} w_m \mathcal{N}(\mathbf{o}; \mu_m, \Sigma_m) \tag{13}$$

where $\mathcal{N}(\mathbf{o}; \mu, \Sigma)$ denotes the evaluation of a normal distribution for observation $\mathbf{o}$ with mean vector μ and covariance matrix Σ. 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 λ are assumed to be diagonal such that $\text{diag}\{\Sigma\} = \sigma$, as substantial benefit can be attained by reducing the number of parameters that need to be estimated.

Given a world model $\lambda_{\mathcal{W}} = \{w_{\mathcal{W}_m}, \mu_{\mathcal{W}_m}, \Sigma_{\mathcal{W}_m}\}_{m=1}^{M}$ 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) [3] which is a form of MAP estimation incorporating the EM-algorithm. Please note for the purposes of clarity we have omitted any indexes to viewpoint.