

# Face Recognition with Image Sets Using Manifold Density Divergence

Ognjen Arandjelović<sup>†</sup> Gregory Shakhnarovich<sup>‡</sup> John Fisher<sup>‡</sup> Roberto Cipolla<sup>†</sup> Trevor Darrell<sup>‡</sup>

<sup>†</sup>Department of Engineering  
University of Cambridge  
Cambridge, CB2 1PZ, UK

<sup>‡</sup>Computer Science and AI Lab  
Massachusetts Institute of Technology  
Cambridge 02139 MA, USA

{oa214,cipolla}@eng.cam.ac.uk {gregory, fisher, trevor}@csail.mit.edu

## Abstract

*In many automatic face recognition applications, a set of a person’s face images is available rather than a single image. In this paper, we describe a novel method for face recognition using image sets. We propose a flexible, semi-parametric model for learning probability densities confined to highly non-linear but intrinsically low-dimensional manifolds. The model leads to a statistical formulation of the recognition problem in terms of minimizing the divergence between densities estimated on these manifolds. The proposed method is evaluated on a large data set, acquired in realistic imaging conditions with severe illumination variation. Our algorithm is shown to match the best and outperform other state-of-the-art algorithms in the literature, achieving 94% recognition rate on average.*

## 1. Introduction

Automatic face recognition (AFR) has long been one of the most active research areas in computer vision. In the last two decades a vast number of different AFR algorithms has been developed – Bayesian eigenfaces [20], Fisherfaces [4], elastic bunch graph matching [18], and the 3D morphable model [8, 23], to name just a few popular ones. These methods have achieved very good accuracy on a small number of controlled test sets.

In sharp contrast is the real-world performance of AFR, which has been, to say the least, disappointing. Even in very controlled imaging conditions, such as those used for passport photographs, the error rate has been reported to be as high as 10% [10], while in less controlled environments the performance degrades even further [9]. We believe that the main reason for this apparent discrepancy between the results reported in the literature and those observed in the real world is that the assumptions that most AFR methods rest upon are hard to satisfy in practice (see Section 2).

Training a system in certain imaging conditions (single illumination, pose and motion pattern) and being able to recognize under arbitrary changes in these conditions can be considered to be the hardest problem formulation for AFR. However, in many practical applications this is too strong a requirement. For example, it is often possible to ask a subject to perform random head motion under varying illumination conditions. It is often not reasonable, however, to request that the user perform a strictly defined motion, assume strictly defined poses or illuminate the face with lights in a specific setup. We therefore assume that the training data available to an AFR system are organized in a database where a *set* of images for each individual represents significant (typical) variability in illumination and pose, but does not exhibit temporal coherence and is not obtained in scripted conditions.

The test data – that is, the input to an AFR system – also often consist of a *set* of images, rather than a single image. For instance, this is the case when the data are extracted from surveillance videos. In such cases the recognition problem can be formulated as taking a set of face images from an unknown individual and finding the best matching set in the database of labelled sets. This is the recognition paradigm we are concerned with in this paper.

We approach the task of recognition with image sets from a statistical perspective, as an instance of the more general task of measuring similarity between two probability density functions that generated two sets of observations. Specifically, we model these densities as Gaussian Mixture Models (GMMs) defined on low-dimensional non-linear manifolds embedded in the image space, and evaluate the similarity between the estimated densities via the Kullback-Leibler divergence. The divergence, which for GMMs cannot be computed in closed form, is efficiently evaluated by a Monte Carlo algorithm.

In the next section, we briefly review relevant literature on face recognition in the context of recognition from image sets and of invariance to illumination and pose changes. We

then introduce our model in Section 3, where we discuss the proposed method for learning and comparing face appearance manifolds. Extensive experimental evaluation of the proposed model and its comparison to state-of-the-art methods are reported in Section 4, followed by discussion of the results and an outline of promising directions for future research.

## 2. Previous Work

Good general reviews of recent AFR literature can be found in [2, 13, 29]. In this section, we focus on AFR literature that deals specifically with recognition from image sets, and with invariance to pose and illumination.

**Recognition across illumination** Illumination invariance is perhaps the most significant challenge for AFR: image differences due to changing illumination may be larger than differences between individuals [1]. Most of the work on recognition under varying illumination has been on recognition from single images. Two of the most influential approaches are the illumination cones of Belhumeur *et al.* [5, 15] and the 3D morphable model of Blanz and Vetter [7]. In [5] the authors showed that the set of images of a convex, Lambertian object, illuminated by an arbitrary number of point light sources at infinity, forms a convex polyhedral cone in the image space with dimension equal to the number of distinct surface normals. In [15], Georghides *et al.* successfully used this result for AFR by reilluminating images of frontal faces. In the 3D morphable model method, parameters of a complex generative model which includes the pose, shape and albedo of a face (assumed to be a Lambertian surface) are recovered in an analysis-by-synthesis fashion.

Both illumination cones and the 3D morphable model have significant shortcomings for practical AFR use. The former approach assumes very accurately registered face images, illuminated from seven to nine different well-posed directions for each head pose. This is difficult to achieve in practical imaging conditions (see the sections that follow for typical image data quality). On the other hand, the 3D morphable model requires nontrivial user intervention (localization of up to seven facial landmarks and the dominant light direction) and has convergence problems in the presence of background clutter or facial occlusions (glasses or facial hair).

**Recognition across pose** Broadly speaking, there are three classes of algorithms that allow for pose invariance. The first, a model-based approach, uses an explicit 2D or 3D model of the face, and attempts to estimate the parameters of the model from the input [8, 18, 23]. This is essentially a view-independent representation.

A second class of algorithms consists of global, parametric models, such as the eigenspace method of Murase and Nayar [21], which use a single parametric, typically linear, subspace estimated from all of the views for all of the objects. In AFR tests, such methods are usually outperformed by methods from the third class, view-based techniques (such as the view-based eigenspaces of Pentland *et al.* [22]), in which a separate subspace is constructed for each pose. View-based algorithms usually require an intermediate step in which the pose of the face is determined, and then recognition is carried out using the estimated view-dependent model.

A common limitation of these methods is that they require a fairly restrictive and labour-intensive training data acquisition protocol, in which a number of fixed views are collected for each subject and appropriately labelled. This is not the case for the method proposed in this paper.

**AFR from image sets** Compared to single-shot recognition, face recognition from image sets is a relatively new area of research. Most of the existing algorithms that deal with multi-image input require image *sequences* and use temporal coherence within the sequence to enforce prior knowledge on likely head movements. In the algorithm of Zhou *et al.* [30] the joint probability distribution of identity and motion is modelled using sequential importance sampling, yielding the recognition decision by marginalization. In [19], Lee *et al.* approximate face manifolds by a finite number of infinite extent subspaces and use temporal information to robustly estimate the operating part of the manifold. Some of these approaches use the “still-to-video” scenario, and do not take full advantage of the sets available for training.

While in some cases temporal information may be useful, we are interested in a more general scenario, in which the images in the set may not be temporally consecutive and in fact may have been collected over an extended period of time and under different conditions. It is often difficult to exploit temporal coherence in such cases. Two previous approaches to this problem are the Mutual Subspace Method (MSM) of Fukui and Yamaguchi [14] and the method of Shakhnarovich *et al.* [24]. These methods propose rather simplistic modelling of face pattern variations, essentially representing the face space as a single linear subspace with a Gaussian density. We believe that this restriction explains the variable results attributed to these methods [24, 27], since it does not capture non-linear variation in appearance due to illumination and pose changes. We propose to overcome this limitation with a more flexible, semi-parametric mixture model presented in the next section.

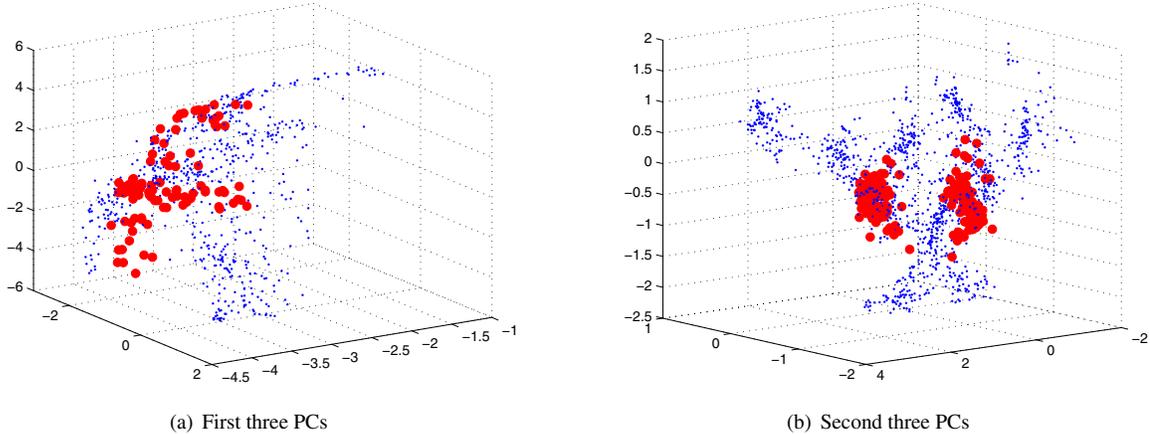


Figure 1. A typical manifold of face images in a training (small blue dots) and a test (large red dots) set. Data used come from the same person and shown projected to the first three (a) and second three (b) principal components. The nonlinearity and smoothness of the manifolds are apparent. Although globally quite dissimilar, the training and test manifolds have locally similar structures.

### 3. Modelling Face Manifold Densities

Under the standard representation of an image as a raster-ordered pixel array, images of a given size can be viewed as points in a Euclidean *image space*. The dimensionality,  $D$ , of this space is equal to the number of pixels. Usually  $D$  is high enough to cause problems associated with the *curse of dimensionality* in learning and estimation algorithms. However, surfaces of faces are mostly smooth and have regular texture, making their appearance quite constrained. As a result, it can be expected that face images are confined to a *face space*, a manifold of lower dimension  $d \ll D$  embedded in the image space [6]. Below, we formalize this notion and propose an algorithm for comparing the estimated densities on the manifolds.

#### 3.1. Manifold Density Model

The assumption of an underlying manifold subject to additive sensor noise leads to the following statistical model: An image  $\mathbf{x}$  of subject  $i$ 's face is drawn from the probability density function (pdf)  $p_F^{(i)}(\mathbf{x})$  within the face space, and embedded in the image space by means of a mapping function  $f^{(i)}: \mathbb{R}^d \rightarrow \mathbb{R}^D$ . The resulting point in the  $D$ -dimensional space is further perturbed by noise drawn from a noise distribution  $p_n$  (note that the noise operates in the image space) to form the observed image  $\mathbf{X}$ . Therefore the distribution of the observed face images of the subject  $i$  is given by:

$$p^{(i)}(\mathbf{X}) = \int p_F^{(i)}(\mathbf{x}) p_n(f^{(i)}(\mathbf{x}) - \mathbf{X}) d\mathbf{x} \quad (1)$$

Note that both the manifold embedding function  $f$  and the density  $p_F$  on the manifold are subject-specific, as denoted by the superscripts, while the noise distribution  $p_n$  is assumed to be common for all subjects. Following accepted practice, we model  $p_n$  by an isotropic, zero-mean Gaussian. Figure 1 shows an example of a face image set projected onto a few principal components estimated from the data, and illustrates the validity of the manifold notion.

Let the training database consist of sets  $S_1, \dots, S_K$ , corresponding to  $K$  individuals.  $S_i$  is assumed to be a set of independent and identically distributed (i.i.d.) observations drawn from  $p^{(i)}$  (1). Similarly, the input set  $S_0$  is assumed to be i.i.d. drawn from the test subject's face image density  $p^{(0)}$ . The recognition task can then be formulated as selecting one among  $K$  hypotheses, the  $k$ -th hypothesis postulating that  $p^{(0)} = p^{(k)}$ . The Neyman-Pearson lemma [12] states that the optimal solution for this task consists of choosing the model under which  $S_0$  has the highest likelihood. Since the underlying densities are unknown, and the number of samples is limited, relying on direct likelihood estimation is problematic. Following [24], we use Kullback-Leibler divergence as a "proxy" for the likelihood statistic needed in this  $K$ -ary hypothesis test.

#### 3.2. Kullback-Leibler Divergence

The Kullback-Leibler (KL) divergence [11] quantifies how well a particular pdf  $q(\mathbf{x})$  describes samples from another pdf  $p(\mathbf{x})$ :

$$D_{KL}(p||q) = \int p(\mathbf{x}) \log \left( \frac{p(\mathbf{x})}{q(\mathbf{x})} \right) d\mathbf{x} \quad (2)$$

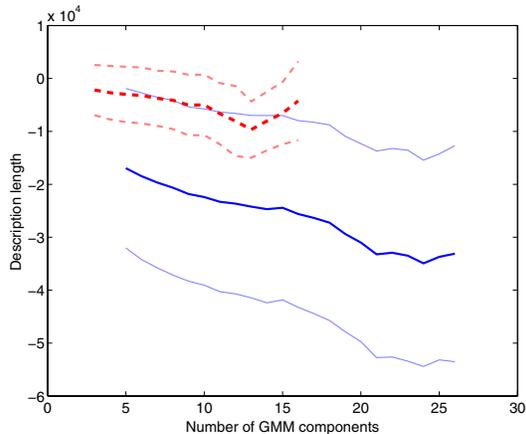


Figure 2. Description lengths for varying numbers of GMM components for training (solid) and test (dashed) sets. The lines show the average plus/minus one standard deviation across sets.

It is nonnegative and equal to zero iff  $p \equiv q$ . Consider the integrand in (2). It can be seen that the regions of the image space with a large contribution to the divergence are those in which  $p(\mathbf{x})$  is significant and  $p(\mathbf{x}) \gg q(\mathbf{x})$ . On the other hand, regions in which  $p(\mathbf{x})$  is small contribute comparatively little. We expect the sets in the training data to be significantly more extensive than the input set, and as a result  $p^{(i)}$  to have broader support than  $p^{(0)}$ . We therefore use  $D_{KL}(p^{(0)}||p^{(i)})$  as a “distance measure” between training and test sets. This expectation is confirmed empirically (see Figure 2). The novel patterns not represented in the training set are heavily penalized, but there is no requirement that all variation seen during training should be present in the novel distribution.

We have formulated recognition in terms of minimizing the divergence between densities on face manifolds. Two problems still remain to be solved. First, since the analytical form for neither the densities nor the embedding functions is known, these must be estimated from the data. Second, the KL divergence between the estimated densities must be evaluated. In the remainder of this section, we describe our solution for these two problems.

### 3.3. Gaussian Mixture Models

Our goal is to estimate the density defined on a complex nonlinear manifold embedded in a high-dimensional image space. As was mentioned in Section 2, global parametric models typically fail to adequately capture such manifolds. We therefore opt for a more flexible mixture model for  $p^{(i)}$ : the Gaussian Mixture Model (GMM). This choice has a number of advantages:

- It is a flexible, semi-parametric model, yet simple enough to allow efficient estimation.

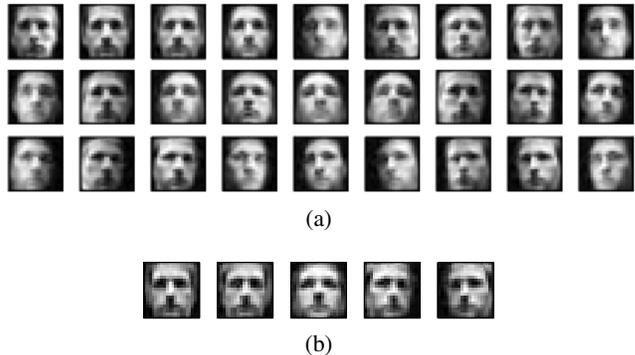


Figure 3. Centres of the MDL GMM approximation to a typical training face manifold, displayed as images (a) (also see Figure 5). These appear to correspond to different pose/illumination combinations. Similarly, centres for a typical face manifold used for recognition are shown in (b). As this manifold corresponds to a video in fixed illumination, the number of Gaussian clusters is much smaller. In this case clusters correspond to different poses only: frontal, looking down, up, left and right.

- The model is generative and offers interpolation and extrapolation of face pattern variation based on local manifold structure.
- Principled model order selection is possible.

The multivariate Gaussian components of a GMM in our method need not be semantic (corresponding to a specific view or illumination) and can be estimated using the Expectation Maximization (EM) algorithm [12]. The EM is initialized by K-means clustering, and constrained to diagonal covariance matrices. As with any mixture model, it is important to select an appropriate number of components in order to allow sufficient flexibility while avoiding overfitting. This can be done in a principled way with the Minimal Description Length (MDL) criterion [3]. Briefly, MDL assigns to a model a cost related to the amount of information necessary to encode the model and the data *given* the model. This cost, known as the description length, is proportional to the likelihood of the training data under that model penalized by the model complexity, measured as the number of free parameters in the model.

Average description lengths for different numbers of components for the data sets used in this paper are shown in Figure 2. Typically, the optimal (in the MDL sense) number of components for a training manifold was found to be 18, while 5 was typical for the manifolds used for recognition. This is illustrated in Figures 3, 4 and 5.

### 3.4. Estimating KL Divergence

Unlike in the case of Gaussian distributions, the KL divergence cannot be computed in a closed form when  $\hat{p}(\mathbf{x})$



Figure 4. Synthetically generated images from a single Gaussian component in a GMM of a training image set. It can be seen that local manifold structure, corresponding to varying head pose in fixed illumination, is well captured.

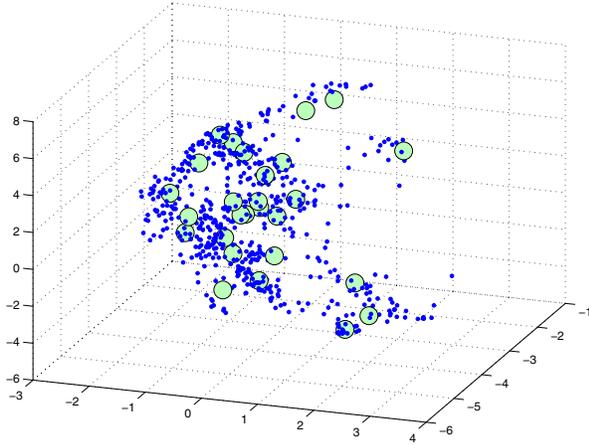


Figure 5. A training face manifold (blue dots) and the centres of Gaussian clusters of the corresponding MDL GMM model of the data (circles), projected on the first three principal components.

and  $\hat{q}(\mathbf{x})$  are GMMs. However, it is straightforward to sample from a GMM. The KL divergence in (2) is the expectation of the log-ratio of the two densities w.r.t. the density  $p$ . According to the law of large numbers [16], this expectation can be evaluated by a Monte-Carlo simulation. Specifically, we can draw a sample  $\mathbf{x}_i$  from the estimated density  $\hat{p}$ , compute the log-ratio of  $\hat{p}$  and  $\hat{q}$ , and average this over  $M$  samples:

$$D_{KL}(\hat{p}||\hat{q}) \approx \frac{1}{M} \sum_{i=1}^M \log \left( \frac{\hat{p}(\mathbf{x}_i)}{\hat{q}(\mathbf{x}_i)} \right) \quad (3)$$

Drawing from  $\hat{p}$  involves selecting a GMM component and then drawing a sample from the corresponding multivariate Gaussian. Figure 4 shows a few examples of samples drawn in this manner. In summary, we use the following approximation for the KL divergence between the test set and the  $k$ -th subject’s training set:

$$D_{KL} \left( \hat{p}^{(0)} || \hat{p}^{(k)} \right) \approx \frac{1}{M} \sum_{i=1}^M \log \left( \frac{\hat{p}^{(0)}(\mathbf{x}_i)}{\hat{p}^{(k)}(\mathbf{x}_i)} \right) \quad (4)$$

In our experiments we used  $M = 1000$  samples.

Age	18-25	26-35	36-45	46-55	65+
Percentage	29%	45%	15%	7%	4%

Table 1. The distribution of ages for the database used in the experiments.



Figure 6. Frames from typical input video sequences used for evaluation of methods in this paper. Notice the presence of cast shadows and drastically varying illumination conditions (different for each frame).

## 4. Empirical Evaluation

Methods in this paper were evaluated on a database with 99 individuals of varying age (see Table 1) and race, and equally represented genders. For each person in the database we collected 7 video sequences of the person in arbitrary motion (significant translation, yaw and pitch, and negligible roll), see Figure 6. Each sequence was recorded in a different illumination setting, at 10 frames per second and  $320 \times 240$  pixel resolution.

The discussion above focused on recognition using fixed-scale face images. A practical AFR system must obtain such images from the available video frames. Before we report the experimental results in Section 4.2, we describe our fully automatic system for extracting and normalizing face image sets from unconstrained video of the subjects. A diagram of the system is shown in Figure 7.

### 4.1. Automatic Acquisition of Face Image Sets

We use the Viola-Jones cascaded detector [26] in order to localize faces in cluttered images. Figure 6 shows examples

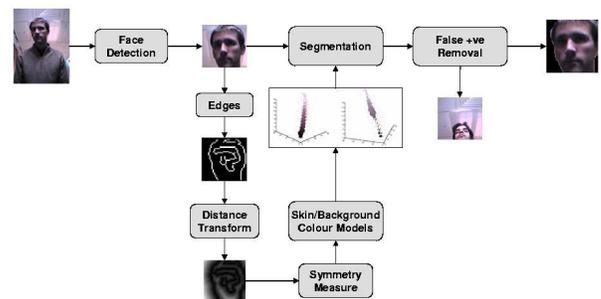


Figure 7. A schematic representation of the face localization and normalization described in Section 4.1.

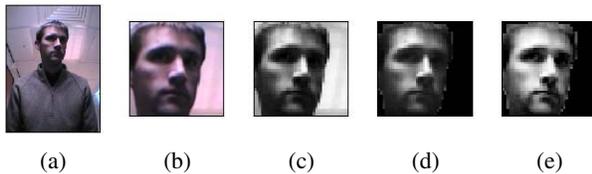


Figure 8. Illustration of the pipeline described in Section 4.1. (a) Original input frame. (b) Face detection. (c) Resizing to the uniform scale of  $40 \times 40$  pixels. (d) Background removal and feathering. (e) The final image after histogram equalization.



Figure 9. Typical false detections identified by our algorithm.

of input frames, and Figure 8 (b) shows an example of a correctly detected face.

**Rejection of false positives** The face detector achieves high true positive rates for our database. A larger problem is caused by false alarms, even a small number of which can affect the density estimates. We use a coarse skin colour classifier to reject many of the false detections. The classifier is based on 3-dimensional colour histograms built for two classes: skin and non-skin pixels [17]. A pixel can then be classified by applying the likelihood ratio test. We apply this classifier and reject detections in which too few ( $< 60\%$ ) or too many ( $> 99\%$ ) pixels are labelled as skin. This step removes the vast majority of non-faces as well as faces with grossly incorrect scales – see Figure 9 for examples of successfully removed false positives.

**Background removal** The bounding box of a detected face typically contains a portion of the background. The removal of the background is beneficial because it can contain significant clutter and also because of the danger of learning to discriminate based on the background, rather than face appearance. This is achieved by *set-specific* skin colour segmentation: Given a set of images from the same subject, we construct colour histograms for that subject’s face pixels and for the near-face background pixels in that set. Note that the classifier here is tuned for the given subject *and* the given background environment, and thus is more “refined” than the coarse classifier used to remove false positives. The face pixels are collected by taking the central portion of the few most symmetric images in the set (assumed to correspond to frontal face images); the background pixels are collected from the 10 pixel-wide strip around the face bounding box provided by the face detector. After classifying each pixel within the bounding box independently, we smooth the result using a simple 2-pass algorithm that en-

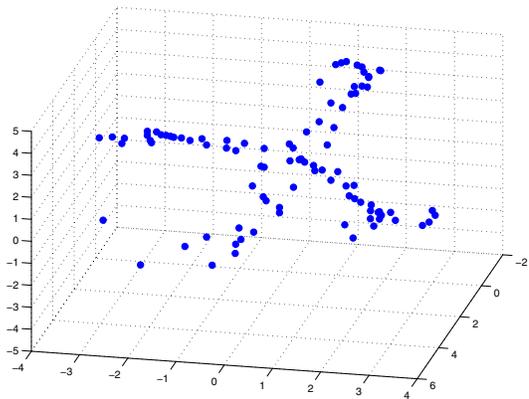


Figure 10. A typical face pose manifold (varying pitch and yaw) acquired in fixed illumination. Four distinct clusters can be seen, corresponding to face looking left, right, up, and down.

forces the connectivity constraint on the face and boundary regions (see Figure 8 (d)).

**Coarse illumination normalization** We normalize for global illumination changes by histogram equalization, performed on face pixels only, after background pixels are removed as described above (see Figure 8 (e)). Additionally, the symmetry of human faces is exploited by augmenting both training and recognition data by their mirror images.

**Pose invariance** Pose variations are typically less problematic than illumination as the corresponding manifold is of lower dimensionality. Figure 10 shows a typical face manifold due to pose changes (pitch and yaw) in an unchanging illumination setup. This manifold, that appears to be 2-dimensional, is accurately reconstructed by our method from components of a GMM, as illustrated by synthetically generated images shown in Figure 4. We therefore do not take any special measures to introduce pose invariance.

## 4.2. Results

We compared the performance of our recognition algorithm to that of:

- The KL divergence-based algorithm of Shakhnarovich et al. (Simple KLD) [24],
- The Mutual Subspace Method (MSM) [28],
- Constrained MSM (CMSM) [14] which projects the data onto a linear subspace before applying MSM,
- Nearest Neighbour (NN) in the set distance sense; that is, achieving  $\min_{\mathbf{x} \in S_0} \min_{\mathbf{y} \in S_i} d(\mathbf{x}, \mathbf{y})$ .

	Proposed method	Simple KLD	MSM	CMSM	Set NN
Ex. 1	96	73	86	96	94
Ex. 2	100	71	92	95	94
Ex. 3	85	63	72	84	79
Mean	94	69	83	92	89
Std	8	5	10	7	9
Sign.		.001	.001	.19	.01

Table 2. Recognition accuracy (%) of the various methods using different training/testing illumination combinations. The last row shows the statistical significance of comparison with the proposed method.

In Simple KLD, we used a principal subspace that captured 90% of the data variance. In MSM, the dimensionality of PCA subspaces was set to 9 [14], with the first three principal angles used for recognition. The constraint subspace dimensionality in CMSM (see [14]) was chosen to be 70. All algorithms were preceded with PCA performed on the entire dataset, which resulted in dimensionality reduction to 150 (while retaining 95% of the variance).

We present three experiments. In each experiment we used all of the sets from one illumination setup as test inputs and the remaining sets as training data. A summary of the experimental results is shown in Table 2. Notice the relatively good performance of the simple NN classifier. This supports our intuition that for training, even random illumination variation coupled with head motion is sufficient for gathering a representative set of samples from the illumination-pose face manifold.

Both MSM-based methods scored relatively well, with CMSM achieving the best performance of all of the algorithms besides the proposed method. That is an interesting result, given that this algorithm has not received significant attention in the AFR community; to the best of our knowledge, this is the first report of CMSM’s performance on a data set of this size, with such illumination and pose variability. On the other hand, the lack of a probabilistic model underlying CMSM may make it somewhat less appealing.

Finally, the performance of the two statistical methods evaluated, the Simple KLD method and the proposed algorithm, are very interesting. The former performed worst, while the latter produced the highest recognition rates out of the methods compared. This suggests several conclusions. Firstly, that the approach to statistical modelling of manifolds of faces is a promising research direction. Secondly, it is confirmed that our flexible GMM-based model captures the modes of the data variation well, producing good generalization results even when the test illumination is not present in the training data set. And lastly, our argument in Section 3 for the choice of the direction of KL divergence is empirically confirmed, as our method performs well even when the subject’s pose is only very loosely controlled.

## 5. Summary and Conclusions

In this paper, we have introduced a new statistical approach to face recognition with image sets. Our main contribution is the formulation of a flexible mixture model that is able to accurately capture the modes of face appearance under broad variation in imaging conditions. The basis of our approach is the semi-parametric estimate of probability densities confined to intrinsically low-dimensional, but highly nonlinear face manifolds embedded in the high-dimensional image space. The proposed recognition algorithm is based on a stochastic approximation of Kullback-Leibler divergence between the estimated densities. Empirical evaluation on a database with 100 subjects has shown that the proposed method, integrated into a practical automatic face recognition system, is successful in recognition across illumination and pose. Its performance was shown to match the best performing state-of-the-art method in the literature and exceed others.

The main direction for future work is to explore the limits of the mixture model and investigate non-parametric approaches. While potentially more expressive, a non-parametric approach poses a number of computational challenges, which are the focus of our current work. Another interesting direction could be to improve the GMM estimation process by using a mixture of probabilistic PCA [25] and thus move away from the current assumption of diagonal covariance. Finally, it may prove beneficial to incorporate more specific domain knowledge, in particular illumination models, in guiding the mixture component estimation.

## Acknowledgements

We would like to thank the Toshiba Corporation, the Cambridge-MIT Institute and DARPA for their kind support for our research, the volunteers from the University of Cambridge Engineering Department whose face videos were entered in our face database, and Trinity College, Cambridge.

## References

- [1] Y. Adini, Y. Moses, and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):721–732, July 1997.
- [2] W. A. Barrett. A survey of face recognition algorithms and testing results. *Systems and Computers*, 1:301–305, 1998.
- [3] A. R. Barron, J. Rissanen, and B. Yu. The Minimum Description Length Principle in Coding and Modeling. *IEEE Transactions on Information Theory*, 44(6):2743–2772, 1998.
- [4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear

- projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.
- [5] P. N. Belhumeur and D. J. Kriegman. What is the set of images of an object under all possible lighting conditions? *In Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 270–277, 1996.
- [6] M. Bichsel and A. P. Pentland. Human face recognition and the face image set’s topology. *Computer Vision, Graphics and Image Processing: Image Understanding*, 59(2):254–261, 1994.
- [7] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. *In Proc. Conference on Computer Graphics and Interactive Techniques*, pages 187–194, 1999.
- [8] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003.
- [9] Boston Globe. Face recognition fails in Boston airport. *Boston Globe*, July 2002.
- [10] British Broadcasting Corporation. Doubts over passport face scans. *BBC News, UK Edition*, October 2004.
- [11] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.
- [12] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., New York, 2nd edition, 2000.
- [13] T. Fromherz, P. Stucki, and M. Bichsel. A survey of face recognition. *MML Technical Report.*, (97.01), 1997.
- [14] K. Fukui and O. Yamaguchi. Face recognition using multi-viewpoint patterns for robot vision. *10th International Symposium of Robotics Research*, 2003.
- [15] A. S. Georghiades, D. J. Kriegman, and P. N. Belhumeur. Illumination cones for recognition under variable lighting: Faces. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1998.
- [16] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes*. Clarendon Press, Oxford, 2nd edition, 1992.
- [17] M. J. Jones and J. M. Rehg. Statistical color models with application to skin detection. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 274–280, 1999.
- [18] B. Kepenekci. *Face Recognition Using Gabor Wavelet Transform*. PhD thesis, The Middle East Technical University, 2001.
- [19] K. Lee, J. Ho, M. Yang, and D. Kriegman. Video-based face recognition using probabilistic appearance manifolds. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 313–320, 2003.
- [20] B. Moghaddam, W. Wahid, and A. Pentland. Beyond eigenfaces - probabilistic matching for face recognition. *In Proc. IEEE Conference on Automatic Face and Gesture Recognition*, pages 30–35, 1998.
- [21] H. Murase and S. Nayar. Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision*, 14:5–24, 1995.
- [22] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 84–91, 1994.
- [23] S. Romdhani, V. Blanz, and T. Vetter. Face identification by fitting a 3D morphable model using linear shape and texture error functions. *In Proc. IEEE European Conference on Computer Vision*, pages 3–19, 2002.
- [24] G. Shakhnarovich, J. W. Fisher, and T. Darrell. Face recognition from long-term observations. *In Proc. IEEE European Conference on Computer Vision*, pages 851–868, 2002.
- [25] M. E. Tipping and C. M. Bishop. Mixtures of probabilistic principal component analyzers. *Neural Computation*, 11(2):443–482, 1999.
- [26] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
- [27] L. Wolf and A. Shashua. Learning over sets using kernel principal angles. *Journal of Machine Learning Research*, 4:913–931, 2003.
- [28] O. Yamaguchi, K. Fukui, and K. Maeda. Face recognition using temporal image sequence. *In Proc. IEEE Conference on Automatic Face and Gesture Recognition*, pages 318–323, 1998.
- [29] W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips. Face recognition: A literature survey. *UMD CFAR Technical Report CAR-TR-948*, 2000.
- [30] S. Zhou, V. Krueger, and R. Chellappa. Probabilistic recognition of human faces from video. *Computer Vision and Image Understanding*, 91(1):214–245, 2003.