

Principal Manifolds and Bayesian Subspaces for Visual Recognition

Baback Moghaddam

TR99-35 December 1999

Abstract

We investigate the use of linear and nonlinear principal manifolds for learning low-dimensional representations for visual recognition. Three techniques: Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Nonlinear PCA (NLPCA) are examined and tested in a visual recognition experiment using a large gallery of facial images from the FERET database. We compare the recognition performance of a nearest-neighbour matching rule with each principal manifold representation to that of a maximum a posteriori (MAP) matching rule using a Bayesian similarity measure derived from probabilistic subspaces and demonstrate the superiority of the latter.

Proceedings of the 7th IEEE International Conference on Computer Vision, ICCV99, September, 1999

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Principal Manifolds and Bayesian Subspaces for Visual Recognition

Baback Moghaddam

TR-99-35 July 1999

Abstract

We investigate the use of linear and nonlinear principal manifolds for learning low-dimensional representations for visual recognition. Three techniques: Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Nonlinear PCA (NLPCA) are examined and tested in a visual recognition experiment using a large gallery of facial images from the “FERET” database. We compare the recognition performance of a nearest-neighbour matching rule with each principal manifold representation to that of a maximum *a posteriori* (MAP) matching rule using a Bayesian similarity measure derived from probabilistic subspaces and demonstrate the superiority of the latter.

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Information Technology Center America; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Information Technology Center America. All rights reserved.

Publication History:-

1. First printing, TR-99-35, July 1999

Principal Manifolds and Bayesian Subspaces for Visual Recognition

Baback Moghaddam
Mitsubishi Electric Research Laboratory
201 Broadway, Cambridge MA 02139, USA
baback@merl.com

Abstract

We investigate the use of linear and nonlinear principal manifolds for learning low-dimensional representations for visual recognition. Three techniques: Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Nonlinear PCA (NLPCA) are examined and tested in a visual recognition experiment using a large gallery of facial images from the “FERET” database. We compare the recognition performance of a nearest-neighbour matching rule with each principal manifold representation to that of a maximum a posteriori (MAP) matching rule using a Bayesian similarity measure derived from probabilistic subspaces and demonstrate the superiority of the latter.

1. Introduction

In recent years, computer vision research has witnessed a growing interest in subspace analysis techniques. In particular, eigenvector decomposition has been shown to be an effective tool for solving problems which use high-dimensional representations of phenomena which are intrinsically low-dimensional. This general analysis framework lends itself to several closely related formulations in object modeling and recognition which employ the *principal modes* or the characteristic *degrees-of-freedom* for description. The identification and parametric representation of data in terms of these “principal manifolds” is at the core of recent advances in parametric descriptions of shape [7], target detection [31, 4, 29], nonlinear image interpolation [3], visual learning [27, 28, 30, 25], automatic face recognition [34, 31, 24] as well as density estimation [25, 26].

Subspace methods also form the basis for exploratory data analysis and pattern recognition where they are used to extract low-dimensional manifolds comprised of statistically uncorrelated or independent variables which tend to simplify tasks such as classification. The Karhunen-Loève

Transform (KLT) [17] and Principal Components Analysis (PCA) [14] are examples of eigenvector-based techniques which are commonly used for dimensionality reduction and feature extraction. Independent Component Analysis (ICA) [6] is yet another linear decomposition which seeks statistically *independent* and non-Gaussian components, modeling the observed data as a linear mixture of (unknown) independent sources. ICA’s proficiency in “blind source separation” [15] has found a particular niche in the analysis of EEG [18] and fMRI [21] signals of the brain. Nonlinear PCA (NLPCA) [16, 8] and nonlinear Principal Surfaces [9, 10] are extensions of these linear techniques. In the following section we will briefly review these principal manifolds, their derivation and subsequent statistical properties. In Section 3, an alternative technique using subspace densities and Bayesian similarity is presented and in Section 4 its performance is compared to Euclidean similarity metrics on principal manifolds.

2. Subspace Representations

Spatiotopic visual data (*e.g.*, images, depth maps, flow fields, *etc.*) can be represented as vectors — *i.e.*, as points in a high-dimensional vector space. For example, a m -by- n pixel 2D image can be mapped to a vector $\mathbf{x} \in \mathcal{R}^{N=mn}$, by lexicographic ordering of the pixel elements.¹ Despite this high-dimensional embedding, the natural constraints of the physical world (and the imaging process) dictate that the data will in fact lie in a lower-dimensional manifold. The primary goal of subspace analysis is to identify, represent and parameterize this manifold in accordance with some optimality criteria. We will now review several leading approaches to obtaining both linear and nonlinear principal manifolds, and highlight their corresponding statistical properties.

¹Without loss of generality we will hereafter assume that the mean image vector $\bar{\mathbf{x}}$ is always subtracted from the data.

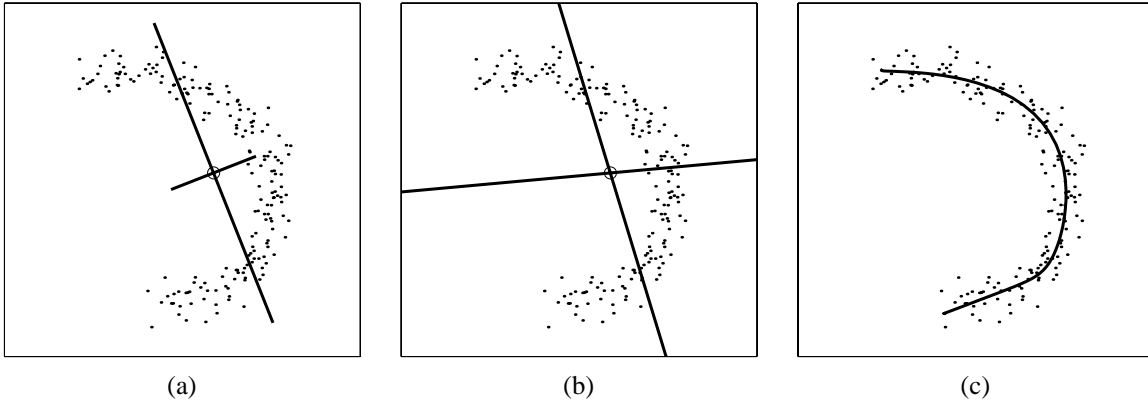


Figure 1. (a) PCA basis (linear, ordered and orthogonal) (b) ICA basis (linear, unordered and non-orthogonal) and (c) Principal Curve (parameterized nonlinear manifold)

2.1. Linear PCA Manifolds

In PCA [14] the basis functions in a discrete Karhunen-Loève Transform (KLT) [17] are obtained by solving the algebraic eigenvalue problem $\mathbf{\Lambda} = \mathbf{\Phi}^T \mathbf{\Sigma} \mathbf{\Phi}$ where $\mathbf{\Sigma}$ is the covariance matrix of the data, $\mathbf{\Phi}$ is the eigenvector matrix of $\mathbf{\Sigma}$ and $\mathbf{\Lambda}$ is the corresponding diagonal matrix of eigenvalues. The unitary matrix $\mathbf{\Phi}$ defines a coordinate transform (rotation) which *decorrelates* the data and makes explicit the *invariant subspace* of the matrix “operator” $\mathbf{\Sigma}$. Most commonly, PCA is a partial KLT which identifies the largest (or principal) eigenvalue eigenvectors for projecting the data: $\mathbf{y} = \mathbf{\Phi}_M^T \mathbf{x}$, where $\mathbf{\Phi}_M$ is a submatrix of $\mathbf{\Phi}$ containing the principal eigenvectors (from here on we will just use $\mathbf{\Phi}$ to denote $\mathbf{\Phi}_M$). PCA can be seen as a linear projection $\mathcal{R}^N \rightarrow \mathcal{R}^M$ onto the lower-dimensional subspace corresponding to the maximal eigenvalues. The main properties of the PCA transform are summarized by the following:

$$\mathbf{x} \approx \mathbf{\Phi} \mathbf{y} \rightarrow \mathbf{\Phi}^T \mathbf{\Phi} = \mathbf{I} \rightarrow E\{y_i y_j\}_{i \neq j} = 0 \quad (1)$$

corresponding to approximate reconstruction, orthonormality of the basis $\mathbf{\Phi}$ and decorrelated principal components, respectively. Figure 1(a) illustrates the PC vectors (columns of $\mathbf{\Phi}$) obtained with a toy data set corresponding to an essentially one-dimensional (nonlinear) manifold. Projection of the data points onto the first PC would then correspond to a 1D linear manifold representation (the 2nd PC, shown as a smaller line segment in the figure, would be discarded in this low-dimensional example).

2.2. Linear ICA Manifolds

Independent Component Analysis (ICA) [15, 6] is similar to PCA except that the components are designed

to be as non-Gaussian as possible (usually by minimizing/maximizing 4th-order cumulants such as kurtosis). ICA is also closely related to “projection pursuit” [12] where maximizing non-Gaussianity promotes statistical *independence*, which is the desired goal. Like PCA, ICA is also a linear projection $\mathcal{R}^N \rightarrow \mathcal{R}^M$ but with different properties:

$$\mathbf{x} \approx \mathbf{A} \mathbf{y} \rightarrow \mathbf{A}^T \mathbf{A} \neq \mathbf{I} \rightarrow P(\mathbf{y}) \approx \prod p(y_i) \quad (2)$$

corresponding to approximate reconstruction, *non-orthogonality* of the basis \mathbf{A} and the near factorization of the joint distribution $P(\mathbf{y})$ into marginal distributions of the (non-Gaussian) ICs. An example of an ICA basis is shown in Figure 1(b) where we see two unordered non-orthogonal IC vectors one of which is roughly aligned with the first PC vector in Figure 1(a) — *i.e.*, the direction of maximum variance. We note that the actual non-Gaussianity and statistical independence achieved in this toy example are minimal at best.

2.3. Nonlinear Principal Manifolds

One of the simplest methods for computing nonlinear principal manifolds is the nonlinear PCA (NLPCA) auto-associative multi-layer neural network [16, 8] shown in Figure 2. Hinton [11] was first to point out that nonlinear networks form useful representations in their hidden layers and Ackley *et al.* [1] were the first to implement an “auto-encoder” trained to reproduce its inputs. The so-called “bottleneck” layer forms a lower-dimensional manifold representation by means of a (weighted-sum-of-sigmoids) nonlinear *projection* function $f(\mathbf{x})$. The resulting PCs \mathbf{y} have an inverse mapping with a similar nonlinear *reconstruction* function $g(\mathbf{y})$, which reproduces the input data as accurately as possible. The defining property of principal

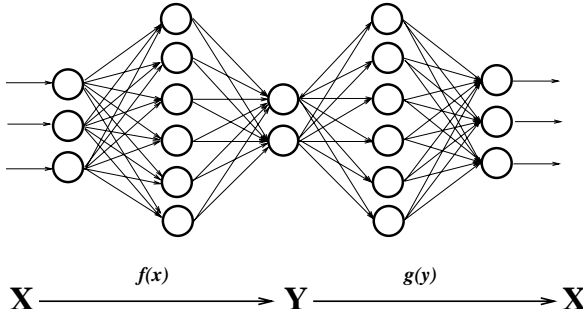


Figure 2. An auto-associative (“bottleneck”) neural network for computing principal manifolds $\mathbf{y} \in \mathcal{R}^M$ in the input space $\mathbf{x} \in \mathcal{R}^N$

manifolds is that the *inverse image* of the manifold in the original space \mathcal{R}^N is (typically) a nonlinear (curved) lower-dimensional surface that “passes through the middle of the data” while minimizing the sum total distance between the data points and their projections on that surface [10]. Note that this is essentially a nonlinear regression on the data. Furthermore, the NLPCA computed by a multi-layer sigmoidal network is equivalent — with certain exceptions² — to a *principal surface* under the more general definition [9, 10]. To summarize, the main properties of NLPCA are:

$$\mathbf{y} = f(\mathbf{x}) \rightarrow \mathbf{x} \approx g(\mathbf{y}) \rightarrow P(\mathbf{y}) = ? \quad (3)$$

corresponding to nonlinear projection, approximate reconstruction and (almost always) no prior knowledge or certainty regarding the joint distribution of the components, respectively. An example of a principal curve is shown in Figure 1(c) which was obtained with a 2-4-1-4-2 layer neural network of the type shown in Figure 2. Note how the principal curve yields a compact and (relatively) accurate representation of the data.

3. Probabilistic Subspaces

The input visual data (or equivalently its manifold representation) can form the basis for simple recognition strategies using Euclidean metrics or normalized correlation. For example, in its simplest form, the similarity measure $S(I_1, I_2)$ between two images I_1 and I_2 (or their manifold projections) can be set to be inversely proportional to the norm $\|I_1 - I_2\|$. Such a simple formulation suffers from a major drawback: it does not exploit knowledge of which types of variation are critical (as opposed to incidental) in expressing similarity. However, one can formulate a

²The class of functions attainable by this type of neural network restricts the projection function $f(\mathbf{x})$ to be smooth and differentiable, hence suboptimal in some cases [19].

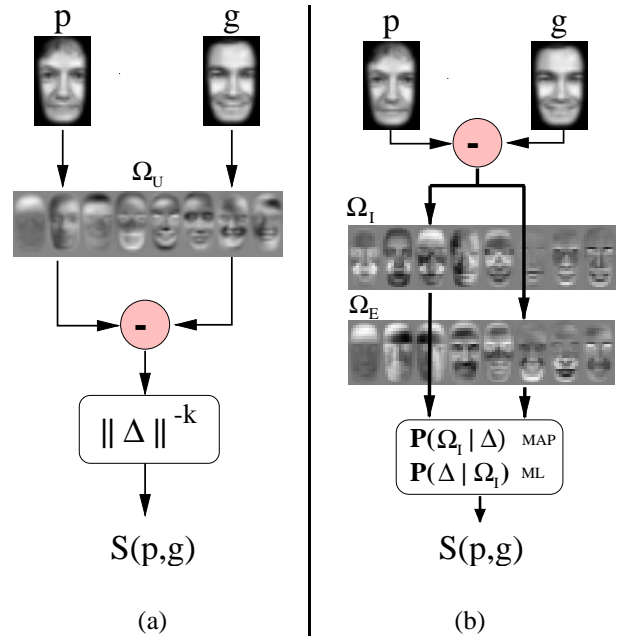


Figure 3. Signal flow diagrams for computing (a) Eigenface similarity and (b) Probabilistic similarity, between two images \mathbf{p} and \mathbf{g} .

probabilistic similarity measure which is based on the probability that the image intensity differences, denoted by $\Delta = I_1 - I_2$, are characteristic of typical variations in appearance of the *same* object. For example, in face recognition, one can define two classes of facial image variations: *intrapersonal* variations Ω_I (corresponding, for example, to different facial expressions of the *same* individual) and *extrapersonal* variations Ω_E (corresponding to variations between *different* individuals). The similarity measure is then expressed in terms of the intrapersonal *a posteriori* probability $S(I_1, I_2) = P(\Omega_I | \Delta)$, by Bayes rule:

$$S = \frac{P(\Delta | \Omega_I) P(\Omega_I)}{P(\Delta | \Omega_I) P(\Omega_I) + P(\Delta | \Omega_E) P(\Omega_E)} \quad (4)$$

The likelihoods $P(\Delta | \Omega_I)$ and $P(\Delta | \Omega_E)$ can be estimated by traditional means (given enough data) or alternatively by the subspace density estimation method [25] using Gaussians or Mixtures-of-Gaussians (for more details see [26]). Furthermore, the priors $P(\Omega)$ can be set to reflect specific operating conditions (*e.g.*, number of test images *vs.* the size of the database) or other sources of *a priori* knowledge regarding the two images being matched. Note that this particular Bayesian formulation casts the standard face recognition task (essentially an \mathbf{m} -ary classification problem for \mathbf{m} objects) into a *binary* pattern classification problem with Ω_I and Ω_E . This simpler problem is then solved using the maximum *a posteriori* (MAP) rule for

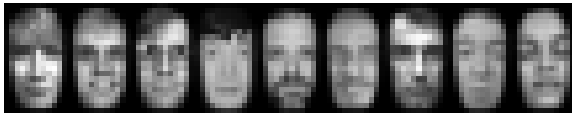


Figure 4. Some representative faces in the dataset.

classification. In other words, two images are determined to belong to the same individual if $P(\Omega_I|\Delta) > P(\Omega_E|\Delta)$, or equivalently, if $S(I_1, I_2) > 0.5$.

Note that this approach requires two linear projections of the difference vector Δ , from which likelihoods can be estimated for the Bayesian similarity measure $S(\Delta)$ as in [26]. Therefore, the projection step is linear while the posterior computation is nonlinear. Because of the double PCA projections required, this approach has been referred to as a “dual eigenspace” technique [23, 22]. This Bayesian method is contrasted to standard PCA or “eigenfaces” in Figure 3. Note the projection of the difference vector Δ onto the “dual subspaces” and the subsequent computation of the posterior in Equation 4. In contrast, the “eigenface” method projects images onto a common (universal) subspace wherein a Euclidean-based similarity is defined.

4. Experiments

Our experimental data consisted of a training “gallery” of 706 individual FERET faces and 1,123 test images or “probes” containing one or more images of every person in the gallery. All these images were aligned and normalized as described in [26]. The multiple probe images reflected different expressions, lighting, glasses on/off, *etc.*. In order to limit the fan-in of the NLPCA network (thus reducing its total number of free parameters) we downsampled the normalized images to 21-by-12 pixels, thus yielding input vectors in a $\mathcal{R}^{N=252}$ space. Examples from our dataset are shown in Figure 4. Note that with $N = 252$ we have nearly 3 times as many training samples than the data dimensionality, thus our parameter estimations (for PCA, ICA, NLPCA and Bayes) were properly over-constrained.

For our recognition experiments we selected a common manifold dimensionality of $M = 20$. This (somewhat arbitrary) choice of M was made for two reasons: it led to a reasonable PCA reconstruction error of $\text{MSE} = 0.0012$ (or 0.12% per pixel with a normalized intensity range of [0,1]) and a baseline PCA recognition rate of $\approx 80\%$ which left a sizeable margin for improvement. To establish fairness in comparisons, all principal manifold projections (PCA, ICA and NLPCA) were required to have the *same* MSE of 0.0012, so that each of them could reconstruct the training set equally well. Naturally, this constraint does not necessarily result in equal recognition rates as we shall see.

4.1. PCA-based Recognition

The baseline algorithm for our face recognition experiments was the standard PCA-based “eigenfaces” [34, 31, 25]. The first 8 principal eigenvectors of our training data are shown in Figure 5 (top). Projection of the training/test set onto this 20-dimensional linear manifold (computed with PCA on the gallery only) and nearest-neighbor matching using a Euclidean metric yielded a 78.98% recognition accuracy with the 1,123 probe images. As a sanity check, we also did full image-vector nearest-neighbor matching (*i.e.*, on $\mathbf{x} \in \mathcal{R}^{252}$) yielding 86.46% (see Figure 6). Clearly, performance is degraded by the $252 \rightarrow 20$ dimensionality reduction, as expected.

4.2. ICA-based Recognition

For recognition experiments with ICA we used two different algorithms based on 4th-order cumulants: the “JADE” algorithm of Cardoso [5] and the fixed-point algorithm of Hyvärinen & Oja [13]. In both algorithms a PCA whitening step (“sphering”) preceded the core ICA decomposition. The corresponding *non-orthogonal* JADE-derived ICA basis is shown in Figure 5 (bottom) — similar basis faces were obtained with Hyvärinen’s method. These basis faces are the columns of the matrix \mathbf{A} in the ICA equation $\mathbf{x} = \mathbf{A}\mathbf{y}$ and their linear combination (specified by the ICs) reconstructs the training data (and also preserves the MSE of the initial PCA step). The ICA manifold projection was obtained using $\mathbf{y} = \mathbf{A}^{-1}\mathbf{x}$ for both training and test images. Nearest-neighbour matching of the ICs, however, gave a recognition rate of 78.90% (with both ICA methods), providing no apparent advantage over PCA. This suggests that seeking non-Gaussian and independent components may not necessarily yield a better manifold representation for *recognition* purposes. We note that the experimental results of Bartlett *et al.* [2] with FERET faces did favor ICA over PCA, but mostly with more difficult time-separated images. Their ICA vs. PCA performance margin at the $\approx 80\%$ recognition level was not as significant.³

4.3. NLPCA-based Recognition

The NLPCA recognition experiments used a fixed neural network architecture (252-64-20-64-252) with logistic sigmoid activation functions. This particular choice of the number of hidden units (64) was based on experimental trials (confirming network convergence) and also using information-theoretic arguments as in [16] such that the

³Compared to Bartlett *et al.* [2] our faces were cropped much tighter, leaving no information regarding hair and face shape and also were much lower in resolution; the combination of these factors makes the recognition task much more difficult.

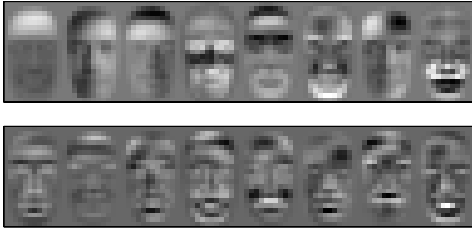


Figure 5. PCA faces (top) and ICA faces (bottom)

total number of free parameters (weights and biases) was roughly one-tenth the total number of data constraints. The nonlinear optimization (on the gallery set) used a conjugate gradient technique with line-search and Polak-Ribiere updates [32] until the $MSE=0.0012$ goal was reached.⁴ Since more than one NLPCA manifold exists for a given MSE, we trained a total of 100 networks, after which both the training and test sets were projected onto the manifold using the projection $f(\mathbf{x})$ and then nearest-neighbour matching of the \mathbf{y} components was performed. The mean recognition rate of the 100 experimental trials was found to be 60.14% with a standard deviation of 8.14% (see Figure 6) and the highest recognition rate obtained in the 100 trials was 74.89%.

4.4. Bayesian Recognition

Bayesian matching requires dual sets of training Δ s for the Ω_I and Ω_E classes. But since we could not form Ω_I vectors from the gallery set (since it contained only one image per person) we divided the total dataset in half such that the new training set contained 353 gallery images (randomly selected) with their corresponding test images (594 probes) which were used to form training samples for both classes, $\{\Delta_i\}_{\Omega_I}$ and $\{\Delta_i\}_{\Omega_E}$. Single Gaussian density estimates were used for the corresponding likelihoods $P(\Delta|\Omega_I)$ and $P(\Delta|\Omega_E)$ using subspace dimensions of $M_I = 10$ and $M_E = 10$ (see [26] for details). Thus the total number of subspace projections required for Bayesian similarity ($M_I + M_E = 20$) was the same as in all the manifold experiments. Finally, using the other (unseen) half of the dataset, *maximum-posteriori* matching of the remaining 529 probes with their corresponding 353 gallery images yielded a 94.71% recognition rate (see Figure 6). Note that the dataset half used for training (density estimation) consisted of entirely different individuals than those used in testing. In contrast, the test set used in the manifold experiments consisted of individuals already represented in the training set. This demonstrates the Bayesian method’s ability to

⁴Note that the usual neural network concerns regarding overfitting and generalization and the preventive use of methods such as “early-stopping” or “weight decay” do not apply here since the final MSE goal is preset.

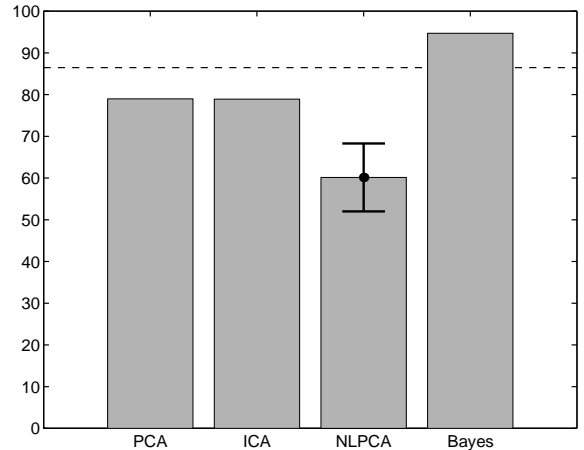


Figure 6. Recognition performance of PCA, ICA, and NLPCA manifolds vs. Bayesian similarity matching (dashed line is the performance of template-matching with the full-dimensional image vectors).

generalize to new data sets, a very desirable property made possible by the probabilistic representation of similarity.

5. Discussion

The relative performance of principal manifolds compared to Bayesian subspace matching is summarized in Figure 6. The advantage of probabilistic matching over Euclidean-based matching with principal manifolds is quite evident ($\approx 15\%$ increase). The NLPCA’s poor performance can be attributed to the general difficulty of computing nonlinear manifolds and the complexity of cost functions riddled with local minima. More sophisticated nonlinear mapping techniques which preserve the local *topology* of the manifold such as [33, 20] are bound to yield better representations for recognition.

Note that both PCA and the dual eigenspaces are uniquely defined for a given training set (thus making experimental comparisons repeatable), whereas NLPCA and ICA are not unique due to the variety of different techniques used to compute them and the stochastic nature of the optimizations involved. Considering the relative computational complexity (of learning), NLPCA required many training epochs and the total number of floating-point operations was significantly large $O(10^{12})$ compared to PCA $O(10^8)$ and ICA $O(10^9)$. Since the Bayesian similarity method’s learning stage requires two separate PCAs, its complexity is essentially twice that of PCA. Considering its significant performance gain and its relative simplicity, the Bayesian subspace method proves to be a very competitive alternative to Euclidean subspace matching methods.

References

- [1] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive Science*, 9(147), 1985.
- [2] M. S. Bartlett, H. M. Lades, and T. J. Sejnowski. Independent component representations for face recognition. In *Proceedings of the SPIE, Vol. 2399: Conference on Human Vision and Electronic Imaging III*, pages 528–539, 1998.
- [3] C. Bregler and S. M. Omohundro. Surface learning with applications to lip reading. In *Advances in Neural Information Processing Systems 6*, pages 43–50, 1994.
- [4] M. C. Burl, U. M. Fayyad, P. Perona, P. Smyth, and M. P. Burl. Automating the hunt for volcanos on venus. In *Proc. IEEE Conf. on Computer Vision & Pattern Recognition*, Seattle, WA, June 1994.
- [5] J-F. Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11(1):157–192, 1999.
- [6] P. Comon. Independent component analysis - a new concept? *Signal Processing*, 36:287–314, 1994.
- [7] T. F. Cootes and C. J. Taylor. Active shape models: Smart snakes. In *Proc. British Machine Vision Conference*, pages 9–18. Springer-Verlag, 1992.
- [8] D. DeMers and G. Cottrell. Nonlinear dimensionality reduction. In *Advances in Neural Information Processing Systems 5*, 1993.
- [9] T. Hastie. *Principal Curves and Surfaces*. PhD thesis, Stanford University, 1984.
- [10] T. Hastie and W. Stuetzle. Principal curves. *Journal of the American Statistical Association*, 84(406):502–516, 1989.
- [11] G. E. Hinton. Learning distributed representations of concepts. In *Proc. Ann. Conf. of the Cognitive Science Society*, volume 1, 1986.
- [12] P. J. Huber. Projection pursuit. *The Annals of Statistics*, 13:435–525, 1985.
- [13] A. Hyvärinen and E. Oja. A family of fixed-point algorithms for independent component analysis. Technical Report A40, Helsinki University of Technology, 1996.
- [14] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, 1986.
- [15] C. Jutten and J. Herault. Blind separation of sources. *Signal Processing*, 24:1–10, 1991.
- [16] M. A. Kramer. Nonlinear principal components analysis using autoassociative neural networks. *AIChE Journal*, 32(2):233–243, 1991.
- [17] M. M. Loève. *Probability Theory*. Van Nostrand, Princeton, 1955.
- [18] S. Makeig, A. J. Bell, T. Jung, and T. J. Sejnowski. Independent component analysis of electroencephalographic data. In *Advances in Neural Information Processing Systems 8*, pages 145–151, 1996.
- [19] E. C. Malthouse. Some theoretical results on nonlinear principal component analysis. Technical report, Northwestern University, 1998.
- [20] T. Martinez and K. Schulten. Topology preserving networks. *Neural Computation*, 7(2), 1994.
- [21] M. J. McKeown, S. Makeig, T. Jung, A. J. Bell, and T. J. Sejnowski. Analysis of fMRI data by blind separation into spatial independent components. *Human Brain Mapping*, 6:160–188, 1998.
- [22] B. Moghaddam, T. Jebara, and A. Pentland. Bayesian modeling of facial similarity. In *Advances in Neural Information Processing Systems 11*, 1998.
- [23] B. Moghaddam, T. Jebara, and A. Pentland. Efficient MAP/ML similarity matching for face recognition. In *Proc. of Int'l Conf. Pattern Recognition*, Brisbane, Australia, August 1998.
- [24] B. Moghaddam, C. Nastar, and A. Pentland. Bayesian face recognition using deformable intensity differences. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, June 1996.
- [25] B. Moghaddam and A. Pentland. Probabilistic visual learning for object detection. In *IEEE Proceedings of the Fifth International Conference on Computer Vision (ICCV'95)*, Cambridge, USA, June 1995.
- [26] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-19(7):696–710, July 1997.
- [27] H. Murase and S. K. Nayar. Learning and recognition of 3D objects from appearance. In *Qualitative Vision Workshop*, New York, June 15-17 1993. CVPR '93.
- [28] H. Murase and S. K. Nayar. Visual learning and recognition of 3D objects from appearance. *International Journal of Computer Vision*, 14(5), 1995.
- [29] S. K. Nayar, S. Baker, and H. Murase. Parametric feature detection. In *Proc. of IEEE Conf. on Computer Vision & Pattern Recognition*, pages 471–477, San Francisco, CA, June 1996.
- [30] S. K. Nayar, H. Murase, and S. A. Nene. General learning algorithm for robot vision. *Neural & Stochastic Methods in Image & Signal Processing*, 2304, 1994.
- [31] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *Proc. of IEEE Conf. on Computer Vision & Pattern Recognition*, Seattle, WA, June 1994.
- [32] L. E. Scales. *Introduction to Non-Linear Optimization*. MacMillan Press, 1985.
- [33] J. Tennenbaum. Mapping a manifold of perceptual observations. In *Advances in Neural Information Processing Systems 10*, 1998.
- [34] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 1991.