1997

Independent components of face images: A representation for face recognition

Marian Stewart Bartlett
University of California San Diego
The Salk Institute
La Jolla, CA 92037
marni@salk.edu

Terrence J. Sejnowski
University of California San Diego
Howard Hughes Medical Institute
The Salk Institute, La Jolla, CA 92037
terry@salk.edu

Abstract

Methods for obtaining representations of face images based on independent component analysis (ICA) are presented. A global ICA representation is compared to a global representation based on principal component analysis (PCA) for recognizing faces across changes in lighting and changes in pose. For each set of face images, a set of statistically independent source images was found through an unsupervised learning algorithm that maximized the mutual information between the input and the output of a nonlinear transformation (Bell & Sejnowski, 1995). These source images comprised the kernels for the representation. The independent component kernels gave superior class discriminability to the principal component kernels. Recognition across changes in pose with the ICA representation was 93%, compared to 87% with a PCA representation, and across changes in lighting ICA gave 100% correct recognition, compared to 90% with PCA.

Introduction

Important advances in face recognition such as "Holons" (Cottrell & Metcalfe, 1991) and "Eigenfaces" (Turk & Pentland 1991) have employed forms of principal component analysis, which addresses only second-order moments of the input (Cottrell & Metcalfe, 1991; Turk & Pentland 1991). Independent component analysis (ICA) is a generalization of principal component analysis (PCA), which decorrelates the higher-order moments of the input (Comon, 1994). In a task such as face recognition, much of the important information is contained in the high-order statistics of the images. A representational basis in which the high-order statistics are decorrelated may be more powerful for face recognition than one in which only the second order statistics are decorrelated, as in PCA representations.

This work examined the statistically independent components of face images in datasets containing changes in lighting and changes in pose. We considered the face images to be a linear mixture of an unknown set of statistically independent source images. The sources were recovered by a matrix of learned filters which produced statistically independent outputs. The independent components were found through an unsupervised learning algorithm that maximized the mutual information between the input and the output of a nonlinear transformation (Bell & Sejnowski, 1995).

We developed and compared three methods for obtaining face representations from the independent components of the image set:

1. The independent components were computed from the full set of N images, producing N independent components as output. A subset of components was selected as kernels for the representation by ordering the sources by the magnitude of the corresponding weights.

2. We next developed a method for separating $M < N$ independent components by performing ICA on a subset of M principal component vectors from the image set. This produced M independent components as kernels for the representation.

3. A third method for selecting a subset of the independent components as kernels was to select the components with the highest between-class to within-class variability. This method gave the best performance, with 93% and 100% correct recognition across pose and lighting respectively for the ICA representation, compared to 86.5% and 96% obtained using the principal component coefficients with the highest between-class to within-class variability.

Image Sets

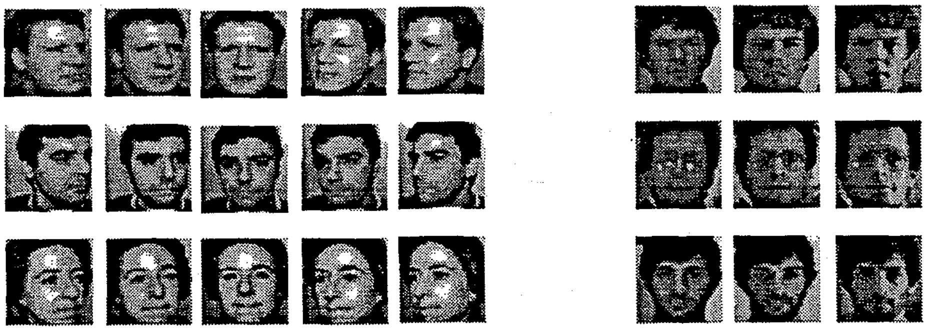

Figure 1: Left: Example images from pose set (Beymer, 1994). Right: Example images from lighting set (Turk & Pentland, 1991).

The analysis was performed on two sets of images, a pose set containing 200 images of forty subjects at five poses each, and a lighting set containing images of sixteen subjects under three different lighting conditions (Figure 1). Eye positions were automatically detected by a template matching algorithm (Beymer, 1994). The faces in the pose set were centered, cropped, and scaled to 60 x 60 by using the eye positions in the frontal pose in each sequence. The lighting images were cropped and scaled using the eye positions in each image. The luminance of the pose set was normalized. The luminance of the lighting set was unaltered. The images were converted to 1x3600 vectors by concatenating the rows, and the mean of each image was shifted to zero.

The Independent Components of Face Images

The images of one image set comprised the rows of a data matrix, X. We considered the face images in X to be a linear mixture of an unknown set of statistically independent source images S, where A is an unknown mixing matrix (Figure 2). The sources were recovered by a matrix of learned filters, W, which produce statistically independent outputs, U.

The weight matrix, W, was found through an unsupervised learning algorithm that maximizes the mutual information between the input and the output of a nonlinear transformation (Bell & Sejnowski, 1995). This algorithm has proven successful for separating randomly mixed auditory signals (the cocktail party problem), and has recently been applied to separating EEG signals (Makeig et al., 1996), fMRI signals (McKeown, in press) and natural scenes (Bell & Sejnowski, in press). The pre-whitening filter in the ICA algorithm has the Mexican-hat shape of retinal ganglion cell receptive fields which remove much of the variability due to lighting (Bell & Sejnowski, in press).

The independent component images contained in the rows of U for the pose dataset are shown in Figure 3 and for the lighting dataset in in Figure 4. The principal components (eigenvectors of the covariance matrix) of the two data sets are also displayed for comparison.

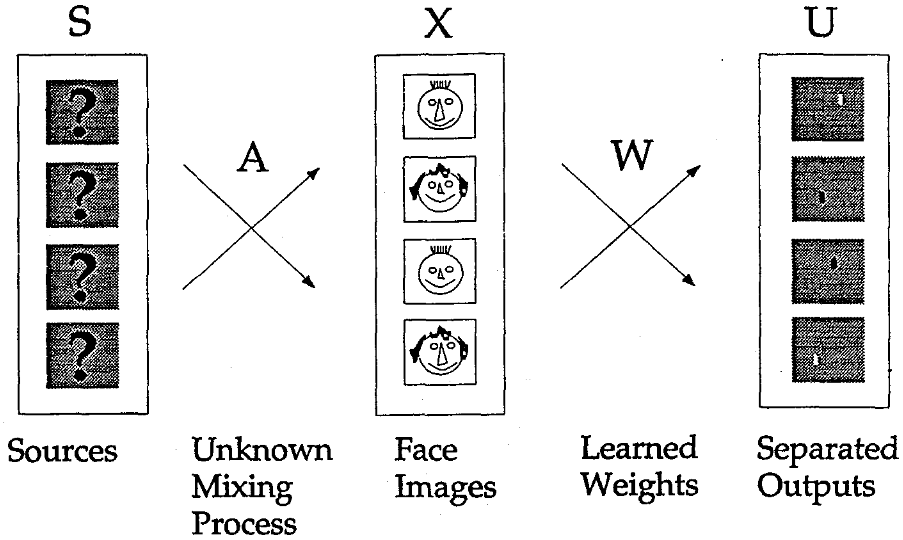

Figure 2: Image synthesis model.

1 A representation from the independent components of the N face images

Performing ICA on the N images in the data set separated N statistically independent source images contained in the rows of U. The rows of the matrix W^{-1} contained the linear combination of source images in U that comprise each face image in X.

$$WX = U \quad \Rightarrow \quad X = W^{-1}U$$

The rows of W^{-1} were chosen as an independent component representation of the face images. An ordering for the ICA representation was provided by the magnitude of the weight vector (row of W) that extracts each source. The magnitude of the weight vector for optimally projecting the source onto the nonlinear transfer function in the ICA algorithm provides a measure of the variance of the original source (Tony Bell, personal communication).

We previously showed (Bartlett & Sejnowski, 1997) that this ICA representation outperformed the PCA representation for recognizing faces across changes in pose (Table 1. The PCA representation of a face consisted of its component coefficients, and was equivalent to the "Eigenface" representation (Turk & Pentland, 1991). A test image was recognized by assigning it the label of the nearest of the other 199 images in Euclidean distance. For the PCA representation, best performance of 85% was obtained with the 120 principal components corresponding to the highest eigenvalues, and for the ICA representation, best performance of 87% was obtained with the 130 independent components with the largest magnitude of the corresponding weight vectors.

	Mutual Information	Percent Correct Recognition
Graylevel Images	.89	.83
PCA	.10	.84
ICA	.007	.87

Table 1: Mean mutual information between all pairs of 10 independent component kernels, 10 principal component kernels, and between 10 original graylevel images from the Pose dataset. Face recognition performance is across all 200 images.

Principal component analysis is optimal for finding a reduced representation that minimizes the reconstruction error, but the kernels that minimize reconstruction error may not be optimal for selecting aspects of the image relevant to classification. The first principal component of the pose image set, for example, appears to be an axis of subject pose and of lighting conditions for the lighting image set (Figure 5). There is a symmetrical relationship between the face images under

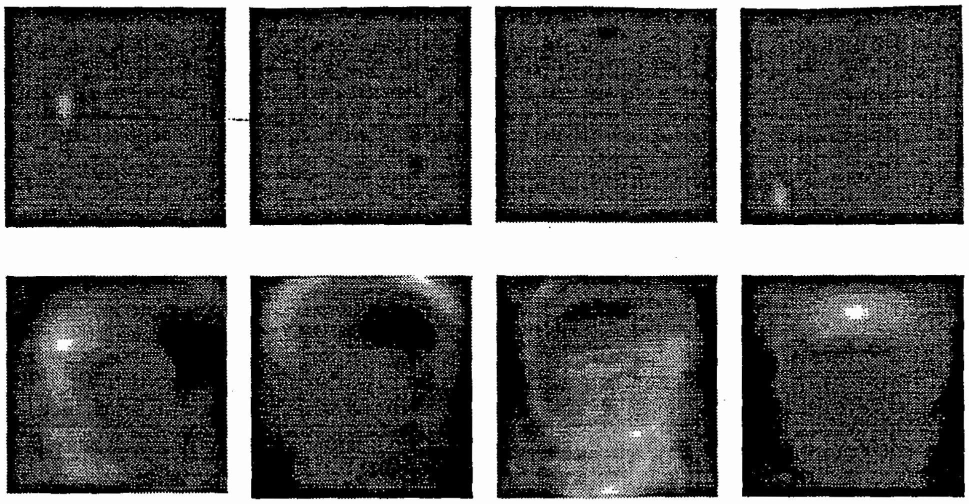

Figure 3: Top: Four independent components of the pose image set. Bottom: First four principal components.

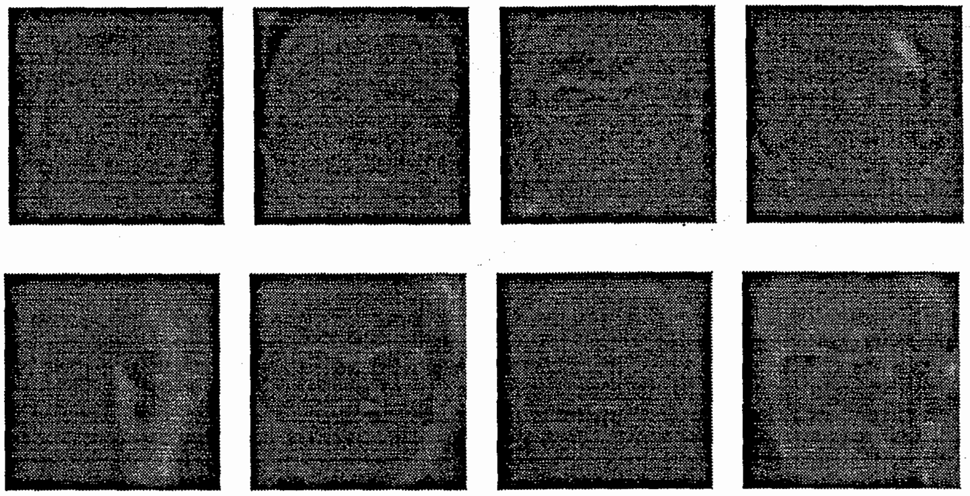

Figure 4: Top: Four independent components of the lighting image set. Bottom: First four principal components.

the different viewing conditions along which the first component axis is oriented, and the sign of the loading indicates the viewing condition. Removing the first principal component from the PCA representation improved performance of the PCA representation to 86%. Recognition performance with the ICA representation likewise improved by removing this component from the data in advance. The procedure described below allowed us to remove this component from the data prior to performing ICA.

2 Separating fewer independent sources: ICA of Eigenfaces

In the procedure described above, the ICA algorithm separated out N independent sources, where N is the number of images. For the pose dataset, $N = 200$. The optimal number of independent sources for recognition performance may be smaller than N. In particular, the pose data set contained 40 subjects + 5 poses = 45 independent parameters and the lighting data set contains 18 subjects + 3 lighting conditions = 18 independent parameters.

In the image synthesis model above, the images in X are assumed to be a linear combination of a set of unknown statistically independent sources. The model is identical if instead of the original

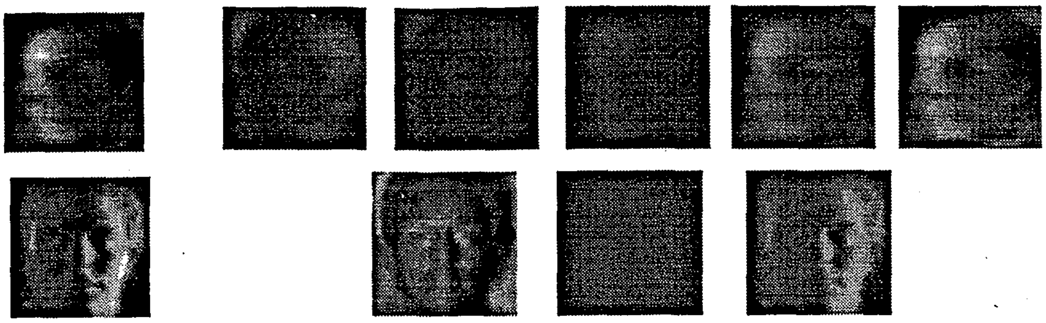

Figure 5: Left: The first principal component of pose data set (top) and of the lighting dataset (bottom). Right: Contribution of the first principal component to the reconstruction of one face at 5 poses (top), and of one face under the three lighting conditions (bottom).

images, X contains some other linear combination of the images. It is possible to obtain a smaller number of sources by using M linear combinations as input to the ICA algorithm, where $M < N$. We chose for these linear combinations a subset of principal component vectors of the data matrix.

Performing ICA on a set of M principal component vectors produced M independent sources, and M corresponding weights. Let U_P denote the Mx3600 matrix of source images obtained from the M principal components. To obtain a representation for all N images in X, we seek an NxM coefficient matrix B such that

$$X = BU_P$$

We solved for B by utilizing the orthogonality of the source matrix U_P and then normalizing the rows of U_P such that $UU^T = I$.

$$XU_P^T = BU_PU_P^T \quad \Rightarrow \quad XU_P^T = B$$

The kernels of our representation were the columns of U_P^T and the coefficients of the representation were contained in the rows of B.

ICA was performed repeatedly on a set of M principal component vectors, where the number of components M ranged from 2 to $N-1$, and classification performance of the ICA representation was compared to the PCA representation using the corresponding M principal component coefficients. The first principal component was excluded. Recognition performance using the ICA representation was consistently equal or superior to that using the original M principal component vectors as kernels. ICA more reliably extracted components that were relevant to class assignment. Although best performance of the two representations was similar (87.5 and 86.5 across pose for the ICA and PCA representations, and 89.6 each across lighting), in the absence of an apriori reason for selecting a specific range of principal components, performance of the ICA representation may be more reliable.

3 Selecting components by class discriminability

Best performance was obtained by selecting the subset of components with the highest ratio of between-class to within-class variability, r:

$$r = \frac{\sigma_{between}}{\sigma_{within}}$$

Where σ_{within} is the sum of the variances within each class and $\sigma_{between}$ is the variance of the class means.

Independent component analysis was performed on principal component vectors $2 - N$, and both the ICA and the PCA components were ordered by the magnitude of r, from largest to smallest. Figure 6 compares the class discriminability ratios for the independent component coefficients and principal component coefficients. The ICA coefficients had consistently greater class discriminability than the PCA coefficients.

Recognition performance of the two representations was compared repeatedly using the M most discriminable components, where the number of components, M, included in the representation ranged from 2 to $N-1$. The ICA representation gave consistently superior recognition performance, regardless of ~~the number of~~ components included. For the ICA representation, best performance for recognition across changes in pose and changes in lighting was 93% and 100% respectively. For the PCA representation, best performance was 86.5% and 89.6%. (See Figure 7). This is analysis of class discriminability is related to Fishers Linear Discriminants (FLD), which is a class specific linear projection to maximize a ratio similar to r. FLD of a principal component representation has been applied to face recognition (Belhumeur et. al. 1996). Our results suggest that FLD of an ICA representation may be even more effective.

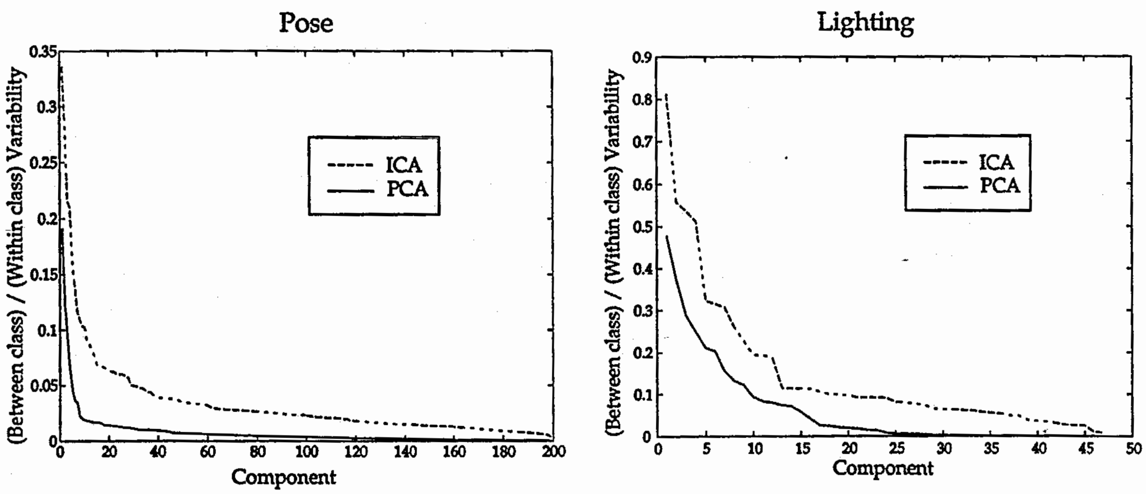

Figure 6: Ratios of between-class to within-class variability (r) for the PCA and ICA representations.

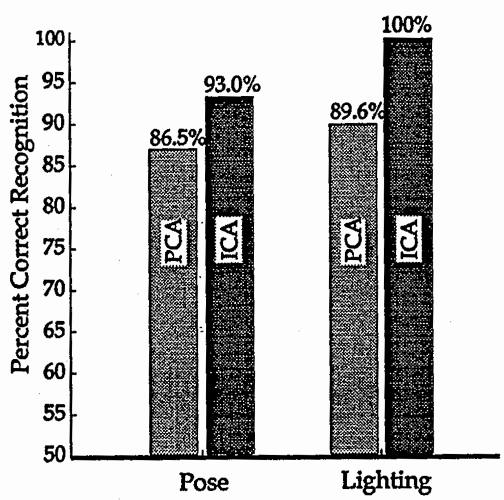

Figure 7: Recognition performance of PCA and ICA representations using the subset of components with the highest between-class to within n-class variability ratio.

Discussion

Independent component representations of face images were derived, and recognition performance across changes in pose and lighting with the ICA representation was compared to performance with a principal component representation. The independent components had greater class discriminability than principal components for recognizing faces across changes in lighting and changes in pose. The ICA representation gave 93% and 100% correct recognition of faces across changes in pose and changes in lighting respectively, compared to 86.5% and 89.6% with a principal component based representation.

The ICA representation of faces presented here is closely related to a method of local feature analysis (LFA), in which the kernels are derived from the pixel-wise correlations of the principal components vectors of an image ensemble (Penev & Atick, 1996). The LFA derivation contains a transform which minimizes the correlations in the output of the kernels.

In LFA, the kernels are optimally matched to the second order statistics of the input ensemble, whereas with the ICA representation, the kernels are optimally matched to the high order statistics of the ensemble as well as the second order statistics. Interestingly, both methods have a tendency to produce local filters, although this constraint was not built into either algorithm.[1]

The independent components of an image ensemble provide a set of statistically independent "features" for coding the images. It has been argued that such a factorial code is advantageous for encoding complex objects that are characterized by high order combinations of features, since the prior probability of any combination of features can be obtained from their individual probabilities (Barlow, 1989; Atick, 1992).

In a task such as face recognition, much of the important information may be contained in the high order spatial relationships in the images. A statistically independent basis set may provide a more powerful representation for face images than principal component based representations such as "Eigenfaces," in which only the second order statistics are decorrelated.

Acknowledgments

We are grateful to Martin McKeown and Michael Gray for helpful discussions on this topic. Support for this work was provided by Lawrence Livermore National Laboratories ISCR agreement B291528, the McDonnell-Pew Center for Cognitive Neuroscience at San Diego, and the Howard Hughes Medical Institute.

References

Atick, J.J. 1992. Could information theory provide an ecological theory of sensory processing? *Network* 3: p. 213-251.

Barlow, H.B. 1989. Unsupervised learning. *Neural Computation* 1: 295-311.

Bartlett, M. Stewart, & Sejnowski, T., 1997. Viewpoint invariant face recognition using independent component analysis and attractor networks. In M. Mozer, M. Jordan, & T. Petsche, Eds., *Advances in Neural Information Processing Systems 9*. Cambridge, MA: MIT Press: 817-823.

Bell, A. & Sejnowski, T., (in press). The independent components of natural scenes are edge filters. *Vision Research.*

Bell, A., & Sejnowski, T., 1995. An information Maximization approach to blind separation and blind deconvolution. *Neural Comp.* 7: 1129-1159.

Belhumeur, P.N., Hespanha, J.P., & Kriegman, D.J. 1996. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projections. *1996 European Conference on Computer Vision.*

Beymer, D. 1994. Face recognition under varying pose. In *Proceedings of the 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* Los Alamitos, CA: IEEE Comput. Soc. Press: 756-61.

Comon, P. 1994. Independent component analysis - a new concept? *Signal Processing* 36:287-314.

Cottrell & Metcalfe, 1991. Face, gender and emotion recognition using Holons. In *Advances in Neural Information Processing Systems 3*, D. Touretzky, (Ed.), Morgan Kaufman, San Mateo, CA: 564 - 571.

[1]The kernels shown in Figure 4 appear more distributed, but this is likely due to the small size of the dataset, and the correspondingly small number of independent components separated. We found in Section 2 that as the algorithm separated larger numbers of independent components, the component images became more local.

Penev, P.S., & Atick, J.J. 1996. Local feature analysis: a general statistical theory for object representation. *Network: Computation in Neural Systems.* 7(3):477-500.

Makeig, S, Bell, AJ, Jung, T-P, and Sejnowski, TJ 1996. Independent component analysis of Electroencephalo~~graphic data, In:~~ *Advances in Neural Information Processing Systems* 8: 145-151.

McKeown MJ, Makeig S, Brown GG, Jung T-P, Kindermann SS, Bell AJ, Sejnowski TJ, (submitted). Analysis of fMRI data by Decomposition into Independent Components.

Turk, M., & Pentland, A. 1991. Eigenfaces for Recognition. *J. Cog. Neurosci.* 3(1):71-86.