# PCA vs. ICA: A comparison on the FERET data set

**Kyungim Baek, Bruce A. Draper, J. Ross Beveridge, Kai She**
Department of Computer Science
Colorado State University
Ft. Collins, CO. 80523, U.S.A.
baek,draper,ross,she@cs.colostate.edu

## Abstract

*Over the last ten years, face recognition has become a specialized applications area within the field of computer vision. Sophisticated commercial systems have been developed that achieve high recognition rates. Although elaborate, many of these systems include a subspace projection step and a nearest neighbor classifier. The goal of this paper is to rigorously compare two subspace projection techniques within the context of a baseline system on the face recognition task. The first technique is principal component analysis (PCA), a well-known "baseline" for projection techniques. The second technique is independent component analysis (ICA), a newer method that produces spatially localized and statistically independent basis vectors. Testing on the FERET data set (and using standard partitions), we find that, when a proper distance metric is used, PCA significantly outperforms ICA on a human face recognition task. This is contrary to previously published results.*

## 1. Introduction

Over the last ten years, face recognition has become a specialized applications area within the field of computer vision. Sophisticated commercial systems have been developed that achieve high recognition rates[1]. Although the details of these systems are generally confidential, most of them include a subspace projection step which projects data from a high-dimensional space to a more meaningful, lower-dimensional space. Face recognition is then implemented as nearest-neighbor classification in this reduced space.

The sophistication of the commercial systems should not be underestimated. These companies have developed methods for pre-processing (and in some cases generating) training data, computing features over images or image differences, and for detecting and registering faces. Unfortunately, the details of commercial systems are generally proprietary. Moreover, many of these techniques are highly specific to face recognition. The underlying subspace projection methods, on the other hand, are more broadly applicable. The goal of this paper is to compare a canonical subspace projection technique – principal component analysis (PCA [1, 2]) – to a newer technique, independent component analysis (ICA [3, 4]).

While PCA has been a popular method in computer vision, especially in face recognition, ICA was originally developed for separating mixed audio signals into independent sources [3]. It is only recently that ICA has been applied to image analysis [5], recognizing faces [6-10] and expressions [11]. Previous results of applying ICA to human face recognition on the FERET database [7, 8] and the Olivetti and Yale databases [10] showed that ICA outperforms PCA, and another report [9] claimed that there is no performance difference between ICA and PCA. In this paper, we make the comparison in the context of a simple, baseline recognition system and the FERET face recognition database. We have tested three different distance metrics – L1 norm, L2 norm, and cosine angle - for both PCA and ICA. We find, contrary to previous reports in the literature, that PCA significantly outperforms ICA when the best performing distance metric is used for each method.

## 2. PCA vs. ICA

PCA is the most widely used subspace projection technique. In PCA, the basis vectors are obtained by solving the algebraic eigenvalue problem $\mathbf{R}^T(\mathbf{X}\mathbf{X}^T)\mathbf{R} = \mathbf{\Lambda}$ where $\mathbf{X}$ is a data matrix whose columns are training samples[2], $\mathbf{R}$ is a matrix of eigenvectors, and $\mathbf{\Lambda}$ is the corresponding diagonal matrix of eigenvalues. The projection of data, $\mathbf{C}_n = \mathbf{R}_n^T\mathbf{X}$, from the original $p$ dimensional space to a subspace spanned by $n$ principal eigenvectors is optimal in the mean squared error sense. That is, the reprojection of

[1] For a comparison of commercial systems, see http://www.dodcounterdrug.com/facialrecognition/FRVT2000/frvt2000.htm

[2] The training samples must be centered, so $\mathbf{X}_i = \mathbf{Y}_i - \mathbf{Y}$, where $\mathbf{Y}_i$ is the i[th] raw training sample and $\mathbf{Y}$ is the mean training sample.

$\mathbf{C}_n$ back into the $p$ dimensional space has minimum reconstruction error. In fact, if $n$ is large enough to include all the eigenvectors with non-zero eigenvalues, the reprojection is lossless.

While the goal in PCA is to minimize the reprojection error from compressed data, the goal of ICA is to minimize the statistical dependence between the basis vectors. Mathematically, this can be written as $\mathbf{WX}^T = \mathbf{U}$, where ICA searches for a linear transformation $\mathbf{W}$ that minimizes the statistical dependence between the rows of $\mathbf{U}$, given a training set $\mathbf{X}$ (as before). Unlike PCA, the basis vectors in ICA are neither orthogonal nor ranked in order. Also, there is no closed form expression to find $\mathbf{W}$. Instead, many iterative algorithms have been proposed based on different search criteria [13]. However, it has been shown that most of the criteria optimized by different ICA algorithms lead to similar or even identical algorithms [14, 15]. In this paper, we will concentrate on InfoMax, one of the best-known ICA algorithms by Bell and Sejnowski [4].

In [7], Bartlett and colleagues first apply PCA to project the data into a subspace of dimension $n$. InfoMax is then applied to the eigenvectors to minimize the statistical dependence among the rows of $\mathbf{U}$ in $\mathbf{WR}_n^T = \mathbf{U}$. It is then possible to reconstruct an approximation to the original images, since $\mathbf{X}^T \approx \mathbf{C}_n^T \mathbf{W}^{-1} \mathbf{U}$. This use of PCA as a pre-processor in a two-step process allows ICA to create subspaces of size $n$ for any $n$. In [8], it is also argued that pre-applying PCA would enhance ICA performance by discarding small trailing eigenvalues before whitening and reduce computational complexity by minimizing pair-wise dependencies. Unfortunately, even with this heuristic ICA basis vectors are much more expensive to compute than PCA basis vectors (on the order of hours rather than seconds).

# 3. The Baseline System

Following the lead of Moon and Phillips [12], we compare PCA and ICA by embedding both algorithms in identical baseline systems. These systems use a subspace projection technique (either PCA or ICA) to compute a set of subspace basis vectors, and then compress a gallery of stored images by projecting them onto the basis vectors. New images are matched to stored images by projecting them onto the basis vectors and matching their projections to the nearest stored (projected) image.

More precisely, the baseline object recognition system used in this study has three steps, and operates on registered face images of three sets: the *training*, *gallery*, and *probe* sets [16]. The first (off-line) step computes the subspace basis vectors from the training images. The second (also off-line) step projects the gallery images into the subspace. The third (on-line) step projects a probe image into the subspace and retrieves the closest galley image to the probe image, as measured in subspace.

Depending on the application, the training and gallery image sets may overlap or even be the same. However, the probe set must always be disjoint from the gallery and training sets.

# 4. Comparing PCA & ICA

We compare PCA and ICA on the task of recognizing faces in the FERET face database. Since face recognition is a significant application, these results are immediately interesting. Moreover, a NIST web site[3] provides results for 10 different algorithms on this task. Because we use their data and replicate their methodology, the ICA numbers presented here can be directly compared to the performance results of these other algorithms. Finally, before the full FERET database was available, there were claims by other researchers that ICA outperforms or equals PCA on face images [7-10]. This paper refutes those claims.

At the same time, readers should be aware that the nature of the recognition task effects the evaluation. In particular, face recognition is more "holistic" in the sense that global properties of the face (e.g. coloring, width, length) are significant. Such features may be more easily captured by PCA than ICA, since ICA basis vectors are more spatially localized than their PCA counterparts (Figure 1). Evaluations on localized recognition tasks, such as recognizing expressions [11], may produce significantly different results.

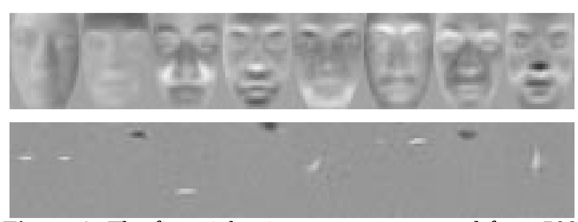

**Figure 1:** The first eight eigenvectors computed from 500 randomly selected images from FERET gallery (top) and eight of 200 ICA basis vectors computed using the technique of [7] (bottom). Note that the eigenvectors are "global" in that they often overlap, assigning significant weights to the same pixels while ICA basis vectors are more spatially localized and never overlap, unlike their PCA counterpart.

## 4.1 The FERET Database

The FERET face recognition database is a set of face images collected by NIST from 1993 to 1997. Each image contains a single face. Prior to processing, the faces are registered to each other, and the backgrounds are eliminated. In this study, only

[3] http://www.itl.nist.gov/iad/humanid/feret/

head-on images are used; faces in profile or at other angles are discarded (although see [6]).

Of particular interest is the structure of the database. The gallery contains 1,196 face images. For this study, the training images are a randomly selected subset of 500 gallery images. Most importantly, there are four different sets of probe images: using the terminology in [12], the *fafb* probe set contains 1,195 images of subjects taken at the same time as the gallery images. The only difference is that the subjects were told to assume a different facial expression than in the gallery image[4]. The *duplicate I* probe set contains 722 images of subjects taken between one minute and 1,031 days after the gallery image was taken. The *duplicate II* probe set is a subset of the duplicate I probe set, where the probe image is taken at least 18 months after the gallery image. The duplicate II set has 234 images. Finally, the *fafc* probe set contains images of subjects under significantly different lighting. This is the hardest probe set, but unfortunately it contains only 194 probe images. In this work, the images were scaled down to 60x50 from the original size of 150x130.

## 4.2 PCA & ICA Implementations

The implementation of PCA used in this comparison is a publicly available C program[5]. It was developed using components from Intel's OpenCV library implementation and compared with other versions of PCA used in the original FERET study. Since the training set contains 500 images, a maximum of 499 non-zero eigenvalues could result. Keeping with standard practice, we keep the first 200 (40%) eigenvectors with the highest eigenvalues.

The implementation of InfoMax is publicly available MATLAB code[6] written by Bell and Sejnowski and used to generate the results in [5] and [7]. InfoMax has several parameters; for this study, the block size was 50, the initial learning rate was 0.001 and, after 1000 iterations, it was reduced every 200 epochs to 0.0005, 0.0002, and 0.0001. We trained InfoMax for 1,600 iterations. These parameters are exactly the same as those used in [7]. To avoid over fitting to the training data, we saved **W** after every 100 iterations, and tested each version. The results presented in the next section are for the best version of **W** on any given probe set.

## 4.3 Results on FERET

The results of comparing PCA and ICA on the FERET data set are given in Table 1. The algorithms are compared by measuring how often a probe image matches the nearest gallery image in subspace. (Alternatively, it is possible to rank them according to how often the true match was one of the K closest images, for any rank K.) Table 1 shows results with three different distance measures for PCA, and cosine angle for ICA[7].

| | PCA | | | ICA |
|---|---|---|---|---|
| | L1 | L2 | cosine | cosine |
| fafb | 80.42 % | 72.80 % | 70.71 % | 78.33 % |
| dup I | 40.30 % | 33.24 % | 35.18 % | 36.15 % |
| dup II | 22.22 % | 14.53 % | 15.38 % | 15.81 % |
| fafc | 20.62 % | 4.64 % | 4.64 % | 6.70 % |

**Table 1:** Correct retrieval rates of PCA and ICA on different probe sets. The probe sets are ordered easiest (fafb) to hardest (fafc).

The most obvious result in Table 1 is that PCA outperforms ICA for every probe set when L1 norm is used. This directly contradicts the previous claims. This is disappointing, since InfoMax is an expensive algorithm that uses PCA as a pre-processor. In effect, the several hours of computation that it took to refine PCA basis vectors into ICA basis vectors reduced performance.

There are several reasons why our results might contradict those reported previously. The most obvious thing would be the distance metric used for comparison. Table 1 clearly shows that L1 norm performs significantly better than both L2 norm and cosine angle, but none of the previous reports used L1 norm for PCA. (In [7], cosine angle was used whereas L2 norm was used in [8] and [9].) We use the distance measure that maximizes the performance of each technique.

Additionally, previous studies were limited to subsets of the FERET database. For example, [7] used 425 training/gallery images and 543 probe images, while [9] used 706 training/gallery images and 1,123 probe images, and [8] used 738 training/gallery images and 369 probes. Table 1, on the other hand, presents results from a larger study with 1,196 gallery images and a total of 2,345 probe images. Moreover, the division of the images into four probe sets and the gallery is the same as has been used for previous FERET evaluation studies.

Since we use the same implementation and parameters as [7], the difference in results (after accounting for the distance metrics) may be attributable to differences in the preprocessing of the input data. We apply NIST's standardization routine to the input images; this was not done in [7].

[4] Subjects were not told what type of facial expression to assume for either the gallery of fafb images, only that the two expressions should be different.
[5] http://www.cs.colostate.edu/evalfacerec/
[6] http://www.cnl.salk.edu/~tony/ica.html

[7] We also tested L1 norm and L2 norm for ICA, and found that cosine angle measure was better for ICA. In this comparison, we only show the best performance by cosine angle for ICA. Cosine angle is also recommended by [7] for ICA.

### 4.4 Statistical Significance

One could question whether the results in Table 1 are statistically significant. The simplest method for determining significance is to model each probe image as a binomial test that either succeeds or fails. McNemar's test can then be used to determine whether one algorithm is significantly better than the other. Under this model, PCA with L1 norm is significantly better than ICA on every probe set. For the fafb and fafc probe sets, the differences are significant to a probability of 99.99% for both. For the duplicate I and duplicate II probe sets, the differences are again significant to 99.96% and 99.87%, respectively.

When L2 norm is used, ICA performs significantly better on the fafb and duplicate I probe sets, but not on the fafc and duplicate II probe sets. Also, there is no difference between ICA and PCA with cosine angle on fafc, duplicate I, and duplicate II probe sets, and ICA is better on fafb probe set. This also contradicts previous reports that show that ICA is better than PCA for all probe sets.

## 5. Conclusion

Previously reported results comparing ICA to PCA for face recognition claim that ICA matches or outperforms PCA [7-10]. Our results with the FERET database contradict this claim. We find that PCA outperforms ICA on all four probe sets when the distance metric for each method is selected to maximize performance. Furthermore, we find that the difference in performance between PCA and ICA is statistically significant.

## 6. Acknowledgements

We thank Marian S. Bartlett for useful conversations and comments. This work was supported in part by DARPA under contract DABT63-00-1-0007.

## 7. References

1. Kirby, M. and L. Sirovich, *Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990. **12**(1): p. 103-107.
2. Turk, M. and A. Pentland, *Eigenfaces for Recognition.* Journal of Cognitive Neuroscience, 1991. **3**(1): p. 71-86.
3. Comon, P., *Independent component analysis; A new concept?* Signal Processing, 1994. **36**(3): p. 287-314.
4. Bell, A.J. and T.J. Sejnowski, *An information-maximization Approach to Blind Separation and Blind Deconvolution.* Neural Computation, 1995. **7**(6): p. 1129-1159.
5. Bell, A.J. and T.J. Sejnowski, *The 'Independent Components' of Natural Scenes are Edge Filters.* Vision Research, 1997. **37**(23): p. 3327-3338.
6. Bartlett, M.S. and T.J. Sejnowski, *Viewpoint invariant face recognition using independent component analysis and attractor networks,* in *Neural Information Processing Systems - Natural and Synthetic*, M. Mozer, M. Jordan, and T. Petsche, Editors. 1997, MIT Press: Cambridge, MA. p. 817-823.
7. Bartlett, M.S., H.M. Lades, and T.J. Sejnowski. *Independent component representations for face recognition*. in *SPIE Symposium on Electronic Imaging: Science and Technology; Conference on Human Vision and Electronic Imaging III*. 1998. San Jose, CA.
8. Liu, C. and H. Wechsler. *Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition*. in *International Conference on Audio and Video Based Biometric Person Authentication*. 1999. Washington, D.C.
9. Moghaddam, B. *Principal Manifolds and Bayesian Subspaces for Visual Recognition*. in *International Conference on Computer Vision*. 1999. Corfu, Greece. p. 1131-1136.
10. Yuen, P.C. and J.H. Lai. *Independent Component Analysis of Face Images.* in *IEEE Workshop on Biologically Motivated Computer Vision.* 2000. Seoul: Springer-Verlag.
11. Donato, G., et al., *Classifying Facial Actions.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999. **21**(10): p. 974-989.
12. Moon, H. and J. Phillips, *Analysis of PCA-based Face Recognition Algorithms,* in *Empirical Evaluation Techniques in Computer Vision,* K. Boyer and J. Phillips, Editors. 1998, IEEE Computer Society Press: Los Alamitos, CA.
13. Hyvärinen, A. and E. Oja, *Independent Component Analysis: Algorithms and Applications.* Neural Networks, 2000. **13**(4-5): p. 411-430.
14. Cardoso, J.-F., *Infomax and Maximum Likelihood for Source Separation.* IEEE Letters on Signal Processing, 1997. **4**: p. 112-114.
15. Hyvärinen, A., *The Fixed-point Algorithm and Maximum Likelihood Estimation for Independent Component Analysis.* Neural Processing Letters, 1999. **10**: p. 1-5.
16. Phillips, P.J., et al., *The FERET Evaluation Methodology for Face-Recognition Algorithms.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000. **22**(10): p. 1090-1104.