

”Equivalence of some common linear feature  
extraction techniques for appearance-based  
object recognition tasks”

M. Asunción Vicente<sup>1</sup>, Patrik O. Hoyer<sup>2</sup>, and Aapo Hyvärinen<sup>2</sup>

<sup>1</sup>Department of Industrial Systems Engineering, Miguel Hernández University, Elche (Alicante), Spain

<sup>2</sup>HIIT Basic Research Unit, Department of Computer Science, University of Helsinki, Finland

## Abstract

Recently, a number of empirical studies have compared the performance of PCA and ICA as feature extraction methods in appearance-based object recognition systems, with mixed and seemingly contradictory results. In this paper, we briefly describe the connection between the two methods and argue that whitened PCA may yield identical results to ICA in some cases. Furthermore, we describe the specific situations in which ICA might significantly improve on PCA.

## Index Terms

Computer vision, object recognition, principal component analysis, independent component analysis.

## I. INTRODUCTION

Over the last few years, there has been increasing interest in appearance-based object recognition due to its successful results in non-controlled scenes compared to classical model-based techniques [7], [8], [11]. Appearance-based recognition approaches are able to manage changes in illumination conditions, shape, pose and reflectance [24] and even to handle translation and partial occlusions [27].

First appearance-based systems found in the literature used Principal Component Analysis (PCA) for dimensionality reduction purposes [17], [24], [25], [27], [30], while recently the use of the independent component analysis (ICA) for feature extraction is preferred by some authors<sup>1</sup>. In fact, a number of empirical studies have claimed that ICA outperforms PCA as a feature extraction method in classification systems [4], [10], [12], [13], [22], [29], [31], [33], although [23],[32] state that both approaches perform equally, [28] suggests that the performance

<sup>1</sup>Although ICA has been connected with sparse representations and edge detection [16], this paper only concerns its application to dimension reduction for subsequent classification.

of ICA is very dependent on the data set, and [3], [14] claim that PCA is superior to ICA. The main goal of this short paper is to explain those seemingly contradictory results and to clarify under which circumstances ICA may outperform PCA.

## II. METHODS FOR DIMENSIONALITY REDUCTION

First of all, it is necessary to have a look at the subtle differences between PCA, Singular Value Decomposition (SVD) and whitening. Although they are intrinsically very close, whitening and PCA transform differ in one important aspect that can affect the performance of the classifier used in the recognition system.

### A. PCA, SVD and whitening

In the original data matrix  $\mathbf{X}$ , each column contains the pixels values of one image vector,  $\mathbf{x}$ , and it is assumed that the data has been centered. The PCA transform of the data matrix  $\mathbf{X}$  of size  $m \times n$  is

$$\mathbf{Y} = \mathbf{U}^T \mathbf{X} \quad (1)$$

where  $\mathbf{U}$  is a  $m \times m$  orthonormal matrix. The principal components (columns of  $\mathbf{U}$ ) are found by recursively seeking out the directions of maximum data variance, under the constraint of orthogonality. The principal component vectors are exactly the eigenvectors of the covariance matrix  $C_{\mathbf{X}} = \frac{1}{N} \mathbf{X} \mathbf{X}^T$  in decreasing order of corresponding eigenvalues. The largest eigenvalue equals the maximal variance, while the corresponding eigenvector determines the direction with the maximal variance. The transformation defined in (1) also gives uncorrelated components.

In (1),  $\mathbf{Y}$  is the original data matrix projected on the PCA subspace defined by the eigenvectors, here, it is possible to reduce the dimension of the data just by selecting a subset of  $k$  eigenvectors from the total set ( $\mathfrak{R}^m \rightarrow \mathfrak{R}^k$ ),

$$\widetilde{\mathbf{Y}} = \widetilde{\mathbf{U}}^T \mathbf{X} \quad (2)$$

We can approach PCA from a slightly different point of view and consider that  $\mathbf{U}$  is one of the matrices of the SVD of the data matrix  $\mathbf{X}$ ,

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \quad (3)$$

where  $\mathbf{U}$  and  $\mathbf{V}$  are  $m \times r$  and  $n \times r$  matrices with orthonormal columns and  $\mathbf{\Lambda}$  is a  $r \times r$  diagonal matrix with the non-negative singular values  $\sigma_j, j = 1, \dots, r$ , arranged in nonincreasing order along the diagonal, and where  $r$  is the rank of  $\mathbf{X}$ .

From (3) it follows that  $\mathbf{X}\mathbf{X}^T\mathbf{U} = \mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}^T$  and  $\mathbf{X}^T\mathbf{X}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}^T\mathbf{\Lambda}$ , demonstrating that the columns of  $\mathbf{U}$  are the eigenvectors of  $\mathbf{X}\mathbf{X}^T$  and the columns of  $\mathbf{V}$  are the eigenvectors of  $\mathbf{X}^T\mathbf{X}$ .

Note that  $\mathbf{X}$  can be written as the sum of  $r$  rank-1 matrices,

$$\mathbf{X} = \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^T \quad (4)$$

This implies that the zero singular values may be ignored since they carry no information. The equation (5) also shows that it is possible to approximate  $\mathbf{X}$  by just using the first  $k$  columns of  $\mathbf{U}$  and  $\mathbf{V}$ ,

$$\mathbf{X} \approx \sum_{j=1}^k \sigma_j \mathbf{u}_j \mathbf{v}_j^T = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^T \quad (5)$$

A zero-mean random vector  $\mathbf{z}$  is said to be *white* if its elements are uncorrelated and have unit variances, this obviously means that their covariance matrix is equal to the unit matrix  $\mathbf{I}$ . As whitening can be accomplished by decorrelation followed by scaling, the PCA technique can be used, and (6) defines the whitening transform of the original data matrix.

$$\widetilde{\mathbf{Z}} = \widetilde{\mathbf{\Lambda}}^{-1}\widetilde{\mathbf{U}}^T\mathbf{X} = \widetilde{\mathbf{V}}^T \quad (6)$$

## B. ICA

ICA [16] tries to explain the original data using statistically independent random vectors. The observed random data matrix  $\mathbf{X}$  is modelled as

$$\mathbf{X} \approx \mathbf{A}\mathbf{S} \quad (7)$$

where  $\mathbf{S}$  is the matrix containing the statistically independent random vectors and  $\mathbf{A}$  is the mixing matrix. We can also write (7) as:

$$\mathbf{S} = \mathbf{W}\mathbf{X} \quad (8)$$

where  $\mathbf{W}$  is equal to the pseudoinverse matrix of  $\mathbf{A}$ , that is  $\mathbf{W} = \mathbf{A}^\dagger$ . Typically, in ICA algorithms, vectors  $\mathbf{w}_i$  are sought such that the rows of  $\mathbf{S}$  have maximally non-gaussian distributions and are mutually (approximately) uncorrelated. A simple way to do this is to first whiten the data as in (6), and then seek orthogonal non-normal projections ( $\mathbf{R}$ ):

$$\mathbf{S} = \mathbf{R}^T \tilde{\mathbf{\Lambda}}^{-1} \tilde{\mathbf{U}}^T \mathbf{X} = \mathbf{R}^T \tilde{\mathbf{Z}} \quad (9)$$

so in (9) it is shown that in fact in this case ICA is a whitening operation followed by a rotation, and the ICA model can be also written as

$$\mathbf{X} \approx \tilde{\mathbf{U}} \tilde{\mathbf{\Lambda}} \tilde{\mathbf{V}}^T = \underbrace{\tilde{\mathbf{U}} \tilde{\mathbf{\Lambda}} \mathbf{R}}_{\mathbf{A}} \underbrace{\mathbf{R}^T \tilde{\mathbf{V}}^T}_{\mathbf{S}} = \mathbf{A}\mathbf{S} \quad (10)$$

Here, we wish to emphasize that in the ICA model neither  $\mathbf{A}$  nor  $\mathbf{W}$  are constrained to be orthogonal. Rather, the constraint (exact [15] or approximate [5], depending on the choice of algorithm) is that the transform be decorrelating, meaning that the rows of  $\mathbf{S}$  are (exactly or approximately) orthogonal. This is because strongly non-orthogonal rows imply strong linear correlations between the estimated components, which is not allowed since the goal was to get

independent components, and independence entails uncorrelatedness. Thus, all ICA algorithms which output approximately uncorrelated components with approximately equal variances are essentially performing an orthogonal transform of the whitened data.

A number of popular ICA algorithms exist. These include FastICA [15], [16], Infomax [5], [18], Comons algorithm [9], and KernelICA [2]. In the case of FastICA and KernelICA, the data is first whitened and subsequently an orthonormal separating matrix is sought, as in (9). In contrast, Infomax does not strictly enforce complete linear decorrelation. Note that this implies that in cases where higher-order dependencies in the data are strong relative to the second-order dependencies, Infomax may yield a decomposition which is not even approximately an orthogonal transform of the whitened data. Comon's algorithm may also give a significantly non-orthogonal transformation from the whitened data due to the different normalization employed, as described later in this section.

In Figures 1 and 2, the results of transforming two-dimensional artificial data sets using PCA, whitening and the ICA model (by means of FastICA (Fig.1) and Extended Infomax [18] (Fig.2)) are illustrated. In the top row of each figure, the original data set has a uniform distribution on a parallelogram (a subgaussian data set); while the original data set in the bottom row has a sparse distribution (a supergaussian data set).

Figure 1 has been drawn using FastICA, so it shows the ICA transformation as in (9). In the top row, the original data set has a uniform distribution on a parallelogram (Figure 1(a)) but the components are not independent and it is possible to predict the value of one of them from the value of the other. The uncorrelated data set is shown in Figure 1(b), the direction with the maximal variance is the vertical axis, and the second principal axis is the horizontal one. The whitened data is shown in Figure 1(c), whitening gives the ICs only up to an orthogonal transformation. And, finally, the independent data set appears in Figure 1(d). The same process is shown in the bottom row, but in this case the original data set has a supergaussian distribution.

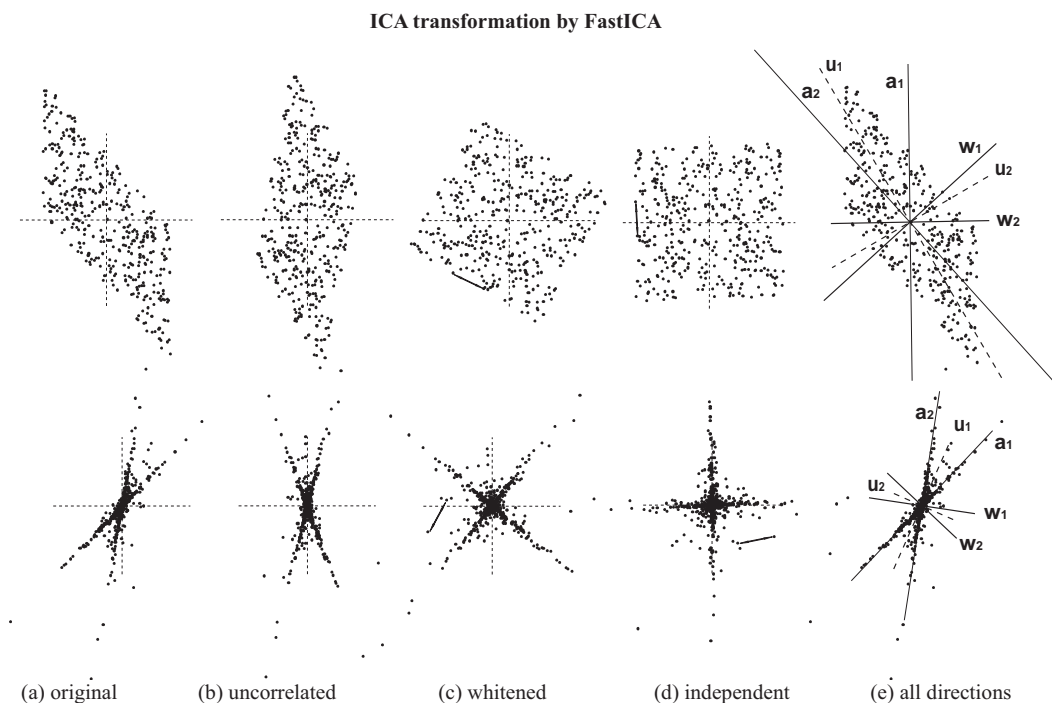


Fig. 1. Two artificial examples: a subgaussian dataset (top row) and a supergaussian dataset (bottom row) (a), both transformed by PCA (b), whitening (c) and using the ICA model by means of FastICA (d). The last figure in each row (e) shows the original dataset with the ICA ( $\mathbf{a}_1, \mathbf{a}_2$  and  $\mathbf{w}_1, \mathbf{w}_2$ ) and PCA directions ( $\mathbf{u}_1, \mathbf{u}_2$ ).

Figure 1(e) shows the original datasets with the ICA and PCA directions. There, we may appreciate that the constraint of orthogonality holds on PCA directions ( $\mathbf{U}$ ) and does not hold on ICA directions ( $\mathbf{W}$ ).

It should be noticed how whitening changes distances between points and so, the distance between two points is not the same in the uncorrelated data and the whitened one. However, the distance between two points is equal in the whitened data and the independent one as the two are equivalent up to a rotation.

In order to show that (for this simple artificial dataset) the IC's obtained by Infomax are also, approximately, an orthonormal transformation of the whitened data, Figure 2 shows the same artificial data sets as Figure 1 but the IC's have been obtained (Fig. 2(b)) by means of the Extended Infomax algorithm [18], which can estimate both sub- and supergaussian components.

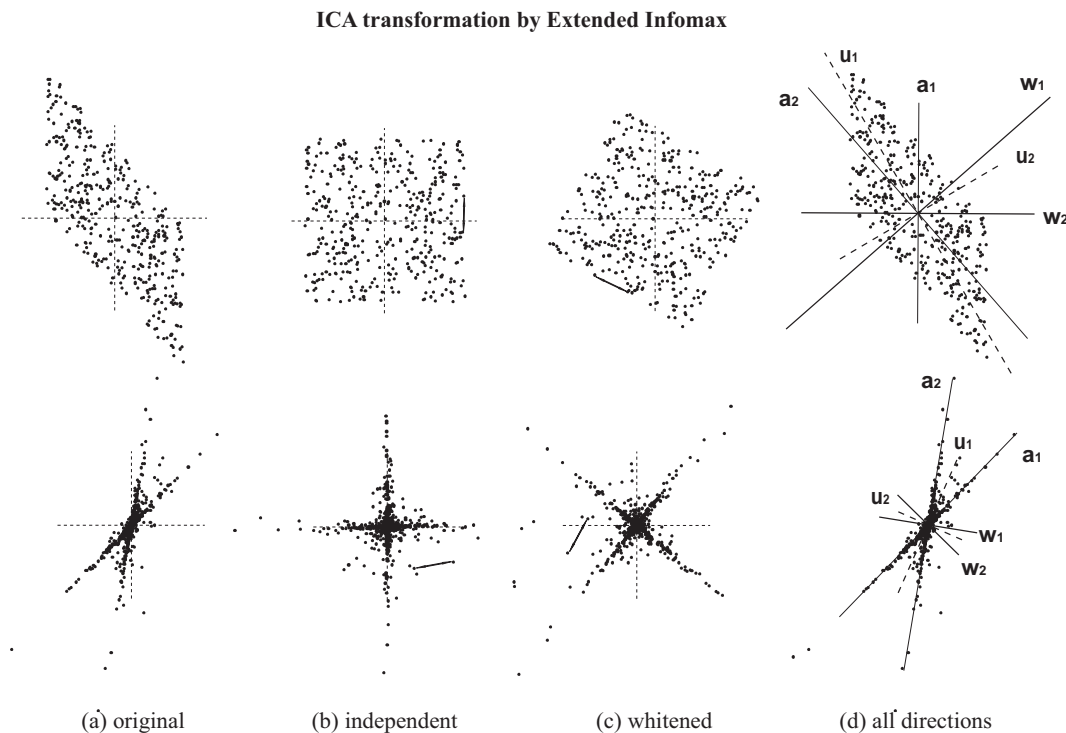


Fig. 2. Two artificial examples: a subgaussian dataset (top row) and a supergaussian dataset (bottom row) (a), both transformed using the ICA model by means of Extended Infomax (b). The whitened data (c), copied from Fig.1 for comparison. The last figure in each file (d) shows the original dataset with the ICA ( $\mathbf{a}_1, \mathbf{a}_2$  and  $\mathbf{w}_1, \mathbf{w}_2$ ) and PCA directions ( $\mathbf{u}_1, \mathbf{u}_2$ ).

It is possible to see that the independent data (Fig. 2(b)) obtained from Infomax is, approximately, a rotated version of the whitened data (Fig. 2(c), copied from Fig.1(c) for comparison purposes). Note, however, that these results cannot always be extended to real-world datasets, where the results of Infomax may not be orthogonal from the whitened data [4]. This fact will be clearly shown on section VI.

Although the above equivalence holds for the most widely used ICA algorithms [2], [5], [15], [16], [18] it does not quite hold for the algorithm of Comon [9]. This is because his algorithm renormalizes the components so that the basis vectors (columns of  $\mathbf{A}$ ) have unit norm. Since this also changes the variance of the independent components (rows of  $\mathbf{S}$ ), the transform is in this case *not* an approximate rotation from the whitened data. Hence, distances, and angles, may



change. However, with a few notable exceptions [19], [20], [21], almost all [3], [4], [10], [12], [13], [14], [22], [23], [28], [29], [31], [32], [33] PCA/ICA comparisons for pattern recognition purposes have used FastICA or Infomax (or its extended version) to perform ICA.

### III. DIFFERENT ARCHITECTURES: INDEPENDENCE IN A OR IN S?

Previous papers [4], [10], [32] have described two different *architectures* for the ICA decomposition of an image set. The difference between these two is a simple choice of where we want the independence. Do we want the *basis images* (columns of  $\mathbf{A}$ ) to be mutually independent, or do we seek a decomposition where the *coefficients* (rows of  $\mathbf{S}$ ) are mutually independent? The former corresponds to architecture I, the latter to architecture II [4], [10].

In equations, the question is whether we write  $\mathbf{X} \approx \widetilde{\mathbf{U}}\widetilde{\mathbf{A}}\mathbf{R}\mathbf{R}^T\widetilde{\mathbf{V}}^T$ , and optimize the orthogonal matrix  $\mathbf{R}$  to make  $\mathbf{S} = \mathbf{R}^T\widetilde{\mathbf{V}}^T$  as independent and non-gaussian as possible, or whether we write  $\mathbf{X} \approx \widetilde{\mathbf{U}}\mathbf{R}\mathbf{R}^T\widetilde{\mathbf{A}}\widetilde{\mathbf{V}}^T$ , and optimize  $\mathbf{R}$  to make  $\mathbf{A} = \widetilde{\mathbf{U}}\mathbf{R}$  as independent and non-gaussian as possible. The concept about the two types of architectures is illustrated in Figure 3.

In either case, the *independent components*, that is the matrix having been optimized for

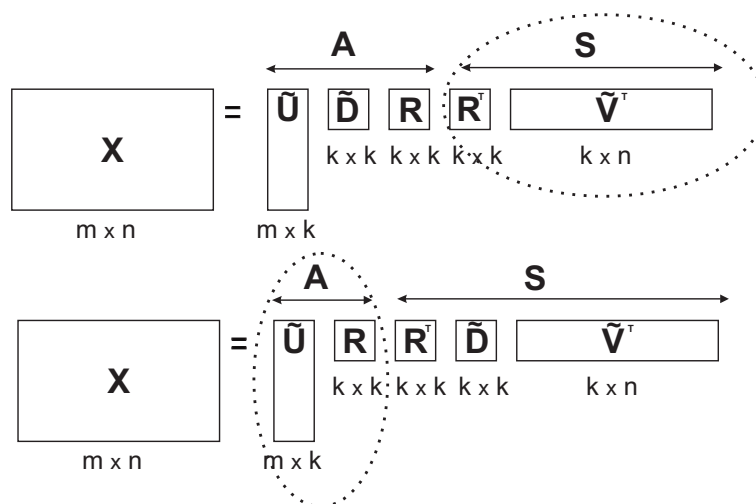


Fig. 3. Different architectures for the ICA decomposition of an image set: do we want independence in A or in S?

independence, is always an orthogonal transformation from the *whitened* data given by SVD, at least in the case of FastICA. As we shall argue in the next section, this implies that if a rotationally-symmetric classifier is subsequently used, there is no point in performing the optimization of  $\mathbf{R}$ .

#### IV. ROTATION INVARIANT CLASSIFIERS

Most common classifiers are invariant to a rotation of the data space. This makes intuitive sense, if one does not have any pre-existing knowledge on the structure of the data space, it seems reasonable to build a classifier which does not care about data rotations. Typical examples are all classifiers based on the Euclidean distances, or based on the angles, between data points [17], [24], [30]. As it was shown in Equation (9), and further argued in Section 3, the independent components are the result of a simple rotation of the whitened data given by SVD. Hence, a rotation-invariant classifier will not do any better nor any worse when fed with the independent components than when fed with the whitened data. Thus, ICA gives you absolutely no advantage over SVD. Empirical verification of this claim is given in Section VI. Note also that, before us, Yang et al [32] came to the exact same conclusion based on experiments on the FERET face database.

However, there are many scenarios in which ICA can give you a different (and hence possibly better) result than SVD. First, if the used classifier is not rotation-invariant, there may be advantages to performing the rotation ICA gives [6]. Second, if a feature selection step is employed and only a subset of all components are used for classification [4], the subspaces selected can differ, and hence the classification results may differ as well. And finally, when using Comon's algorithm, or in some cases when using Infomax, the found components are not necessarily a rotation of the whitened data and hence the equivalence can disappear.

In conclusion, when employing ICA it is important to use either feature selection or a rotation

variant classifier, or both, since otherwise there is little justification for performing an expensive ICA optimization, instead of simply employing the SVD.

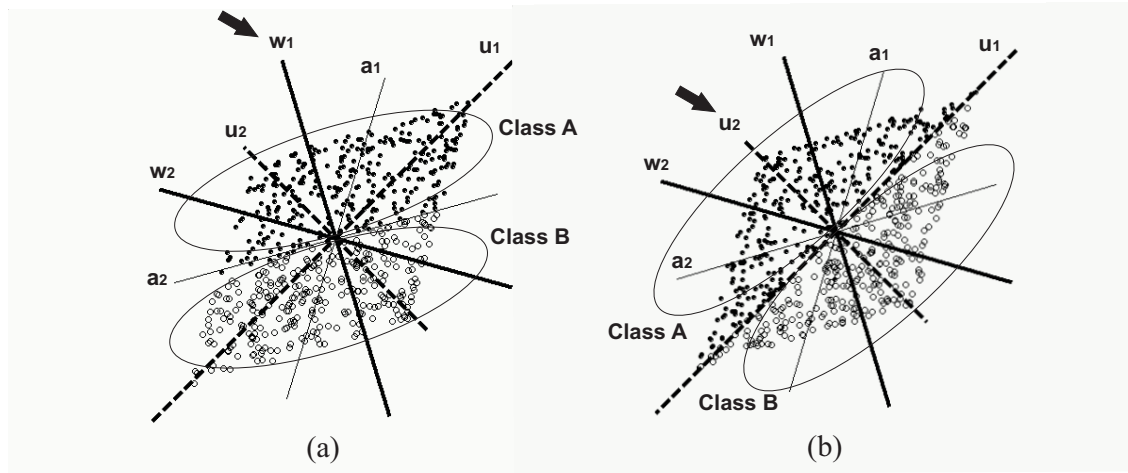


Fig. 4. The *best* selection of the features may be different depending on the classes of data set. In Fig. 4(a) the classes are perfectly separated using just the projection over the ICA direction  $w_1$ , while in Fig. 4(b), the classes are better separated using the direction of the eigenvector  $u_2$

## V. SELECTING SUBSETS OF COMPONENTS

In the previous sections, we showed that, as typically applied, there is no reason to prefer ICA over whitened data. However, the performance of ICA and PCA may differ when a subset of components is used for classification. To illustrate this fact, an artificial data set similar to the subgaussian data in Figure 1 has been assigned to two different classes in two extreme examples in Figure 4. In both plots, the directions of the eigenvectors,  $U$ , and the ICA directions,  $W$  and  $A$ , from the artificial data set are drawn. As both techniques are unsupervised, these directions are independent from the classes in the data set, so they are equal for both extreme examples. In Figure 4(a) it is easy to appreciate how the classes are perfectly separated using just the projection over the ICA direction  $w_1$ , while in Figure 4(b), the classes are better separated using the direction of eigenvector  $u_2$ . Therefore, in each example, a feature selection step may help to

reduce the dimensionality and improve the classification. If no feature selection is carried out, ICA and whitened PCA perform exactly equally well on both datasets, provided a rotationally symmetric classifier is used.

## VI. EXPERIMENTAL RESULTS

Some experiments with real data have been carried out in order to test our hypothesis about the ICA/whitened PCA equivalence. Both the ORL face database [1] and the COIL-100 object database [26] have been used, and the results have been obtained both for FastICA and Infomax algorithms. The classifier used is a 1-NN with Euclidean distance. Due to space limitations, only a summary of the results obtained is shown in this paper; the full set of experiments and the Matlab code used are available at [http://isa.umh.es/arvc/personal/suni/ica\\_pca/index.html](http://isa.umh.es/arvc/personal/suni/ica_pca/index.html).

Figure 5 (experiment 12 of the above mentioned url) shows the results obtained by whitened PCA, FastICA and Infomax with the ORL face database. For a number of components ranging from 5 to 200, the average and standard deviation of 10 repetitions of the same experiment are shown. No feature selection is performed.

From these results, it becomes clear that FastICA and whitened PCA perform equally, and that Infomax shows some differences. These differences are more clear when the number of components used is either very low or very high. Particularly, when the number of components is very high, the recognition rates obtained by Infomax are clearly better than those of FastICA or whitened PCA. Although these good results are found when the number of components is far from the optimum in terms of recognition rates, they deserve a further study in order to find an explanation, which may be related to the non-orthogonality of the components returned by Infomax. Such study is left for further work, as it falls beyond the scope of this short paper.

Figure 6 shows similar experiments to those of figure 5 but considering feature selection (experiment 21 of the above mentioned url). In these figures, the original data (40 classes with

10 examples per class, so the size of  $\mathbf{X}$  is  $10304 \times 400$ ) has been reduced to just 5, 10 and 20 components ( $5 \times 400$ ,  $10 \times 400$  and  $20 \times 400$ ) by means of whitening, Infomax and FastICA. For each method, the best component is selected as the one which offers the best classification results when used alone. The same process is performed to find the best subset of 2 and 3 components using an exhaustive search or "brute force" method, so we have computed the recognition rates of all possible subsets of 2 and 3 components. In this way, the results show the recognition rates that could be obtained with each algorithm if a perfect selection of components is carried out. We have used a quite small number of components in order to be able to perform an exhaustive search. Such search quickly becomes computationally intractable when the number of components grows.

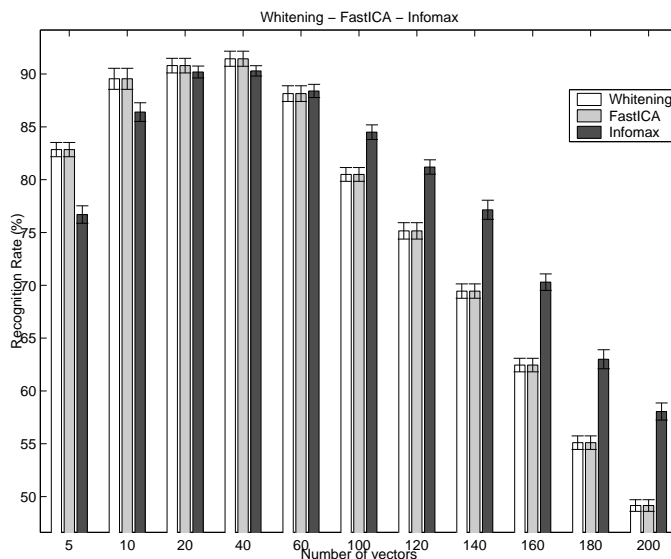


Fig. 5. ICA/whitened PCA comparison using the ORL face database using all components.

The results show clearly that a feature selection process may have a strong influence in the recognition rates obtained by each algorithm. FastICA and whitened PCA no longer perform equally; and the same applies to Infomax. Eventhough the results of the three algorithms differ to a great extent, there is no clear winner. As all the algorithms are unsupervised, the results

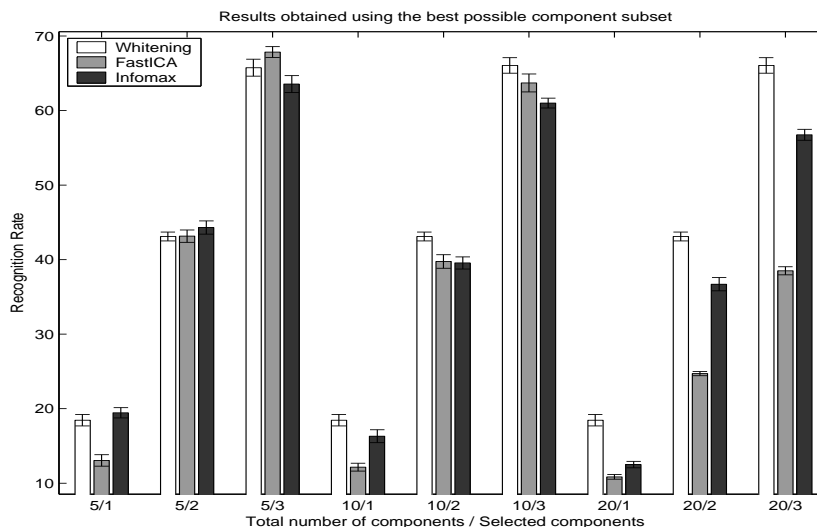


Fig. 6. ICA/whitened PCA comparison using the ORL database when an exhaustive feature selection process is carried out.

are highly dependent on the class distributions. This fact was previously illustrated in section V with the artificial datasets from figure 4.

## VII. DISCUSSION

Visual appearance-based object recognition methods are usually based on feature extraction techniques such as PCA and ICA. In the present paper it has been shown how ICA and PCA are closely connected and under which circumstances their performance is totally equivalent. Their performance may differ significantly if (1) a feature selection process is carried out, (2) a non-rotationally-invariant classifier is used, (3) a renormalization such as that used in Comon's algorithm is performed, or (4) Infomax is used and the data yields independent components which are not close to an orthogonal transform from the whitened data.

## VIII. ACKNOWLEDGEMENTS

The authors wish to thank the anonymous reviewers for their insightful suggestions during the review process. P.O.H. was supported by the Academy of Finland, project #204826. M.A.V.

would also like to thank C. Fernández for helpful discussions and collaboration.

## REFERENCES

- [1] <http://www.cl.cam.ac.uk/Research/DTG/attarchive/facedatabase.html>, *AT&T Laboratories Cambridge*
- [2] F.R. Bach and M.I. Jordan, "Kernel Independent Component Analysis", *JMLR*, 3(Jul):1-48, Mar. 2002.
- [3] K. Baek, B.A. Draper, and J.R. Beveridge, K. She, "PCA vs. ICA: A comparison on the FERET data set", *Proc. 6th Joint Conference on Information Sciences*, pp.824-827, Mar. 2002.
- [4] M.S. Bartlett, J.R. Movellan, and T. J. Sejnowski, "Face Recognition by Independent Component Analysis", *IEEE. Trans. on Neural Networks*, vol. 13, no. 6, pp. 1450-1464, Nov. 2002.
- [5] A.J. Bell and T.J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution", *Neural Computation*, vol.7, pp. 1129-1159, 1995.
- [6] M. Bressan and J. Vitria, "On the Selection and Classification of Independent Features", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1312-1317, 2003.
- [7] R. Brunelli and T. Poggio, "Face recognition: features versus templates", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042-1052, 1993.
- [8] R.J. Campbell and P.J. Flynn, "A Survey Of Free-Form Object Representation and Recognition Techniques", *Computer Vision and Image Understanding*, vol. 81, pp. 166-210, 2001.
- [9] P. Comon, "Independent component analysis, a new concept? ", *Signal Processing*, vol. 36, no.3, pp. 287-314, 1994.
- [10] B. Draper, K. Baek, M.S. Bartlett, and R. Beveridge, "Recognizing Faces with PCA and ICA", *Comput. Vis. Image Underst.*, vol. 91, no.1-2, pp. 115-137, Jul. 2003.
- [11] S. Edelman, "Computational theories of object recognition", *Trends in Cognitive Sciences*, vol. 1, pp. 296-304, 1997.
- [12] H.K. Ekenel, and B. Sankur, "Multiresolution face recognition", *Image and Vision Computing*, vol. 23, pp. 469-477, 2005.
- [13] J. Fortuna, D. Shuurman, and D. Capson, "A comparison of PCA and ICA for object recognition under varying illumination", *Proc. 16th International Conference on Pattern Recognition*, vol. 3, pp. 11-15, 2002.
- [14] J. Fortuna, P. Quick, and D. Capson, "A comparison of subspace methods for accurate position measurement", *Proc. 6th IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 16-20, Mar. 2004.
- [15] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis", *IEEE Trans. on Neural Networks*, vol. 10, no.3, pp. 626-634, 1999.
- [16] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.
- [17] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no.1, pp. 103-108, Jan. 1990.

- [18] T.W. Lee, M. Girolami and T.J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources", *Neural Computation*, vol. 11, no. 2, pp. 417-441, 1999.
- [19] C. Liu and H. Wechsler, "Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition", *Second International Conference on Audio- and Video-based Biometric Person Authentication, AVBPA'99, Washington D. C.*, March 22-24, 1999.
- [20] C. Liu and H. Wechsler, "Independent Component Analysis of Gabor Features for Face Recognition", *IEEE Trans. Neural Networks*, vol. 14, no. 4, pp. 919-928, 2003.
- [21] C. Liu, "Enhanced Independent Component Analysis and Its Application to Content Based Face Image Retrieval", *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, no. 2, pp. 1117-1127, 2004.
- [22] T. Martiriggiano, M.Leo, T. D'Orazio and A. Distanti "Face Recognition by Kernel Independent Component Analysis", *Lecture Notes in Artificial Intelligence*, vol. 3533, pp. 55-58, 2005.
- [23] B. Moghaddam, "Principal Manifolds and Probabilistic Subspaces for Visual Recognition", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 1-10, Jun. 2002.
- [24] H. Murase and S.K. Nayar, "Visual Learning and Recognition of 3D Objects from Appearance", *Int'l J. Computer Vision*, vol. 14, pp. 5-24, 1995.
- [25] S.K. Nayar, S.A. Nene and H. Murase, "Subspace Methods for Robot Vision", *IEEE Trans. Robotics and Automation*, vol. 12, no.5, pp. 750-758, 1996.
- [26] S.A. Nene, S.K. Nayar and H. Murase, "Columbia Object Image Library (COIL-100)", *Tech. Report CU-CS-006-96*, 1996.
- [27] K. Ohba and K. Ikeuchi, "Detectability, uniqueness, and reliability of eigen windows for stable verification of partially occluded objects", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 9, pp. 1043-1048, 1997.
- [28] H.S. Sahambi and K. Khorasani, "A neural-network appearance-based 3-D object recognition using independent component analysis", *IEEE Transactions on Neural Networks*, vol. 14, no.1, pp. 138- 149, 2003.
- [29] D.A. Socolinsky and A. Selinger, "A Comparative Analysis of Face Recognition Performance with Visible and Thermal Infrared Imagery", *Proceedings of the 16th International Conference on Pattern Recognition*, 2002.
- [30] M. Turk and A. Pentland, "Eigenfaces for Recognition", *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.
- [31] M. Vicente, C. Fernández, O. Reinoso and L. Payá, "3D Object Recognition from Appearance: PCA Versus ICA Approaches", *Lecture Notes in Computer Science*, vol. 3211, pp. 547-555, 2004.
- [32] J. Yang, D. Zhang and J. Yang, "Is ICA significantly better than PCA for face recognition?", *Proc. IEEE International Conference on Computer Vision*, pp. 198-203, 2005.
- [33] P.C. Yuen and J.H. Lai, "Independent component analysis of face images", *Proc. IEEE Workshop on Biologically Motivated Computer Vision*, pp. 545-553, 2000.