Experiments on Lattice Independent Component Analysis for Face Recognition

Ion Marqués, Manuel Graña

Computational Intelligence Group, Dept. CCIA, UPV/EHU, Apdo. 649, 20080 San Sebastian, Spain www.ehu.es/ccwintco

Abstract. In previous works we have proposed Lattice Independent Component Analysis (LICA) for a variety of image processing tasks. The first step of LICA is to identify strong lattice independent components from the data. The set of strong lattice independent vector are used for linear unmixing of the data, obtaining a vector of abundance coefficients. In this paper we propose to use the resulting abundance values as features for clasification, specifically for face recognition. We report results on two well known benchmark databases.

1 Introduction

Face recognition [3] is one of the most relevant applications of image analysis. It's a true challenge to build an automated system which equals human ability to recognize faces. There are many different industrial applications interested in it, most of them somehow related to security. Face recognition may consist in the authentication of a user, which a binary decision, or in the identification of a user which is a (large) multiclass problem.

Images of faces, represented as high-dimensional pixel arrays, often belong to a manifold of lower dimension. In statistical learning approaches, each image is viewed as a point (vector) in a d-dimensional space. The dimensionality of these data is too high. Therefore, the goal is to choose and apply the right statistical tool for extraction and analysis of the underlying manifold. These tools must define the embedded face space in the image space and extract the basis functions from the face space. This would permit patterns belonging to different classes to occupy disjoint and compacted regions in the feature space. Consequently, we would be able to define a line, curve, plane or hyperplane that separates faces belonging to different classes. The classical approach applied Principal Component Analysis (PCA) for feature extraction [19], other approaches use the variations of the Linear Discriminant Analysis (LDA) [11,22,21,10,13,20,14,2], or the Locality Preserving Projections (LPP) [7]. Other successful statistic tools include Bayesian networks [12], bi-dimensional regression [9], generative models [8], and ensemble-based and other boosting methods [11].

In this paper we report experimental results with a novel feature extraction method based on the notion of lattice independence: Lattice Independent Component Analysis (LICA) [4]. Lattice independent vectors are affine independent and define a convex polytope. LICA aims to find a set of such lattice independent vectors from the data whose associated convex polytope covers all or most of the data. Feature extraction then consists in the computation of the unmixing process relative to these vectors, which is equivalent to the computation of the convex coordinates relative to them. We explore the performance of this feature extraction process for face recognition over to well known benchmark databases, comparing with Principal Component Analysis (PCA) and Independent Component Analysis (ICA) applied as alternative feature extraction processes.

The paper is organized as follows: Section 2 introduces the LICA approach. Section 3 reports the experimental results. Section 4 gives our conclusions and further work directions.

2 Lattice Independent Component Analysis (LICA)

Lattice Independent Component Analysis is based on the Lattice Independence discovered when dealing with noise robustness in Morphological Associative Memories [16]. Works on finding lattice independent sources (aka endmembers) for linear unmixing started on hyperspectral image processing [6,17]. Since then, it has been also proposed for functional MRI analysis [5] among other.

Under the Linear Mixing Model (LMM) the design matrix is composed of endmembers which define a convex region covering the measured data. The linear coefficients are known as fractional abundance coefficients that give the contribution of each endmember to the observed data:

$$\mathbf{y} = \sum_{i=1}^{M} a_i \mathbf{s}_i + \mathbf{w} = \mathbf{S}\mathbf{a} + \mathbf{w}, \tag{1}$$

where $\mathbf{y}$ is the d-dimension measured vector, $\mathbf{S}$ is the $d \times M$ matrix whose columns are the d-dimension endmembers $\mathbf{s}_i, i = 1, .., M$, $\mathbf{a}$ is the M-dimension abundance vector, and $\mathbf{w}$ is the d-dimension additive observation noise vector. Under this generative model, two constraints on the abundance coefficients hold. First, to be physically meaningful, all abundance coefficients must be non-negative $a_i \geq 0, i = 1, .., M$, because the negative contribution is not possible in the physical sense. Second, to account for the entire composition, they must be fully additive $\sum_{i=1}^{M} a_i = 1$. As a side effect, there is a saturation condition $a_i \leq 1, i = 1, .., M$, because no isolate endmember can account for more than the observed material. From a geometrical point of view, these restrictions mean that we expect the endmembers in $\mathbf{S}$ to be an Affine Independent set of points, and that the convex region defined by them covers *all* the data points.

The *Lattice Independent Component Analysis* (LICA) approach assumes the LMM as expressed in equation 1. Moreover, the equivalence between Affine Independence and Strong Lattice Independence [15] is used to induce from the data the endmembers that compose the matrix $\mathbf{S}$. Briefly, LICA consists of two steps:

Algorithm 1 One step of the cross-validation of LICA for face recognition

1. Build a training face image matrix $X_{TR} = \{\mathbf{x}_j; j = 1, \ldots, m\} \in \mathbb{R}^{N \times m}$. The testing image matrix is denoted $X_{TE} = \{\mathbf{x}_j; j = 1, \ldots, m/3\} \in \mathbb{R}^{N \times m/3}$.
2. Data preprocessing approaches:
   (a) either perform PCA over X, obtaining $T = \{\mathbf{t}_j; j = 1, \ldots, m\} \in \mathbb{R}^{m \times m}$
   (b) or directly do $T = X_{TR}$.
3. Obtain a set of k endmembers using an EIA over T: $E = \{\mathbf{e}_j; j = 1, \ldots, k\}$ from T. Varying EIA parameters will give different E matrices. The algorithm has been testing with α values ranging from 0 to 10.
4. Unmix train and test data: $Y_{TR} = E^{\#} X_{TR}^T$ and $Y_{TE} = E^{\#} X_{TE}^T$.
5. Nearest Neighbor classification: For each image vector $\mathbf{y}_j \in Y_{TE}$
   (a) calculate the Euclidean distance to each training image $\mathbf{v}_j \in Y_{TR}$.
   (b) assign the class to which $\mathbf{y}_j$ belongs as the class of the nearest $\mathbf{v}_j$.
6. Compute performance statistics: classification accuracy

1. Use an Endmember Induction Algorithm (EIA) to induce from the data a set of Strongly Lattice Independent vectors. In our works we use the algorithm described in [6,5]. These vectors are taken as a set of affine independent vectors that forms the matrix $\mathbf{S}$ of equation 1.
2. Apply the Full Constrained Least Squares estimation to obtain the abundance vector according to the conditions for LMM.

The advantages of this approach are (1) that we are not imposing statistical assumptions to find the sources, (2) that the algorithm is one-pass and very fast because it only uses lattice operators and addition, (3) that it is unsupervised and incremental, and (4) that it can be tuned to detect the number of endmembers by adjusting a noise-filtering related parameter. When $M \ll d$ the computation of the abundance coefficients can be interpreted as a dimension reduction transformation, or a feature extraction process.

2.1 LICA for face recognition

Our input is a matrix of face images in the form of column vectors. The induced SLI vectors (endmembers) are selected face images which define the convex polytope covering the data. A face image is defined as a $A_{a \times b}$ matrix composed by $a \cdot b = N$ pixels. Images are stored like row-vectors. Therefore, column-wise the dataset is denoted by $Y = \{\mathbf{y}_j; j = 1, \ldots, N\} \in \mathbb{R}^{n \times N}$, where each $\mathbf{y}_j$ is a pixel vector. Firstly, the set of SLI $X = \{\mathbf{x}_1\} \in \mathbb{R}^{n \times K}$ is initialized with the maximum norm pixel (vector) in the input dataset Y. We chose to use the maximum norm vector as it showed experimentally to be the most successful approach.

We have tested LICA over the original data and over the PCA transformation coefficients. For the PCA we retain all non-null eigenvalue eigenvectors. The maximum number of such eigenvectors is the size of the data sample, because we have much less data samples than the space dimensionality. The classification method performed was a 30 times executed 4-fold cross-validation, randomizing the folds on each iteration; and selecting by euclidean distance the nearest neighbor to decide the class. One step of the cross-validation process is specified in algorithm 1. In this algorithm $E^{\#}$ denotes the pseudo-inverse of the matrix E. Note that we compute the feature extraction process over the training data for each repetition of the data partition into train and test subsamples. When testing PCA as a feature extraction, we retain the eigenvectors with greatest eigenvalues. The algorithm for endmember induction, the EIA, used is the one in [6] which has tolerance parameter α controlling the amount of endmembers detected. In the ensuing experiments we have varied this parameter in order to obtain varying numbers of endmembers on the same data. In other words, in step 3 of algorithm 1 there is implicit an iteration varying the values of α in order to obtain diverse dimensionality reductions.

Fig. 1. An instance of the first 5 eigenfaces (PCA), independent components (ICA) and endmembers (LICA)

3 Experimental results

The recognition task was performed over the ORL database[18] and the Yalefaces database [1]. We did not perform any image registration or spatial normalization. Neither we did perform any face detection process. Images were taken as given from the databases. On Yalefaces we tested a simple normalization consisting in extracting the mean intensity value of the image to all the pixels (to obtain a zero mean) and adding them the middle value of the gray scale interval. Tests covered dimensionality reduction up to 30 components. For ICA and PCA that was accomplished selecting the desired sources and eigenvectors, respectively. For LICA that implies varying the value of the α parameter and observing the number of endmembers detected. Graphic 5 contains the endmembers obtained depending on the α value. Graphic 4 illustrates the relation between α and hit-rate. Table 1 contains the best cross-validation results obtained for each database and feature extraction process. On the ORL database, LICA obtained better results on the original images than on the result of PCA transformation. LICA improves on ICA, with a greater dimensionality reduction. LICA best result is worse than PCA's on this database. For the Yalefaces, the ICA performs better than the other two and LICA improves over PCA. The normalization of the images introduces some improvement in ICA and LICA based approaches, but not in PCA.

Method	prep. data	ORL		Yalefaces original		Yalefaces normalized	
		Acc.	Dim.	Acc.	Dim.	Acc.	Dim.
PCA	-	0.94	25	0.70	25	0.70	27
ICA	PCA	0.86	30	0.76	26	0.80	27
LICA	PCA	0.87	24	0.73	10	0.76	30
LICA	-	0.91	15			0.78	30

Table 1. Face recognition results.

For a better assessment of the algorithm's performance, we show the plots of the recognition accuracy versus the final dimension of the transformed data. These plots represent the average accuracy obtained from the cross-validation repetitions at such dimension reductions. Figure 2 shows the accuracy versus dimension reduction on the ORL database. It can be appreciated that LICA features computed over the original images improve for all dimension over the ICA features and is close to the PCA features. The LICA features computed on the PCA transformed data perform worse than the other approaches for almost all dimensions tested. Figure 3 shows the accuracy versus dimension on the Yalefaces database after the normalization of the images described above. It can be appreciated that PCA performs better for some low dimension but is improved by ICA as the number of dimensions increase. The LICA features on the original images improve steadily with the dimensions approaching the performance of ICA. It's noticeable the good performance obtained over Yalefaces database, taking into account that it includes great illumination variations.

4 Conclusions

We have applied LICA and two well know dimension reduction procedures to feature extraction for face recognition on two well known databases. The results

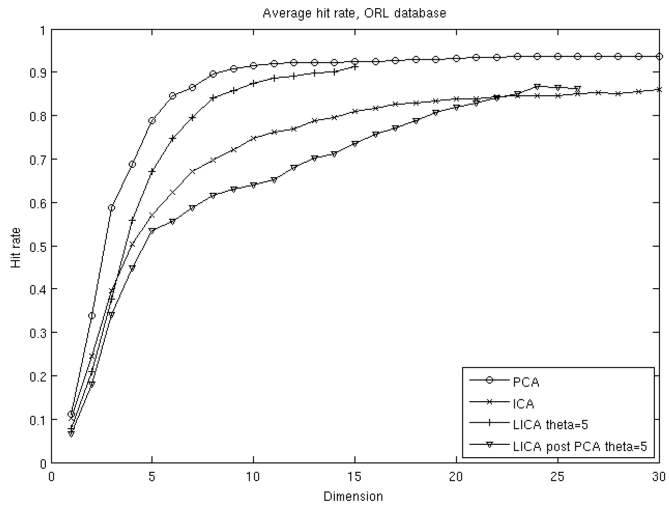

Fig. 2. Plots of accuracy versus dimension on the ORL database

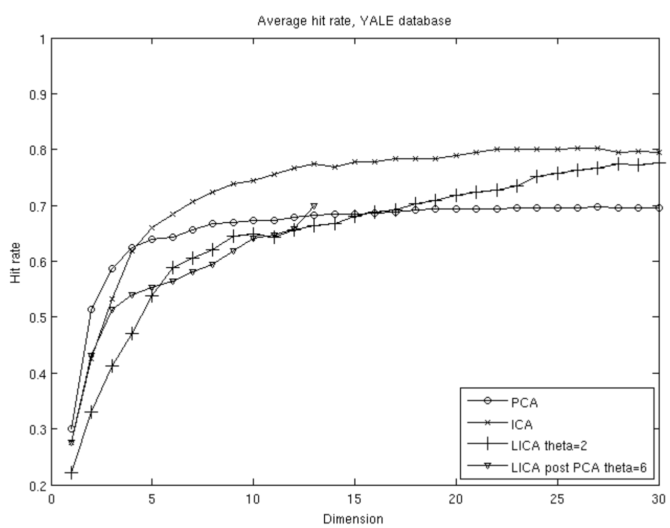

Fig. 3. Plots of accuracy versus dimension on the Yalefaces database

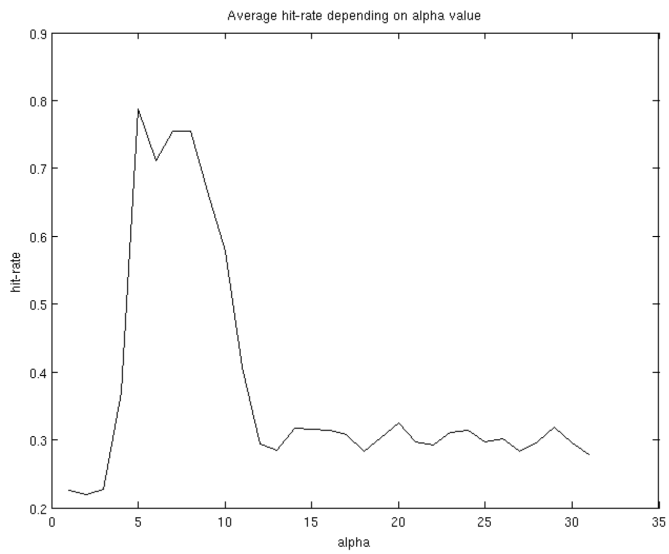

Fig. 4. Accuracy of LICA on the Yalefaces database for different α values.

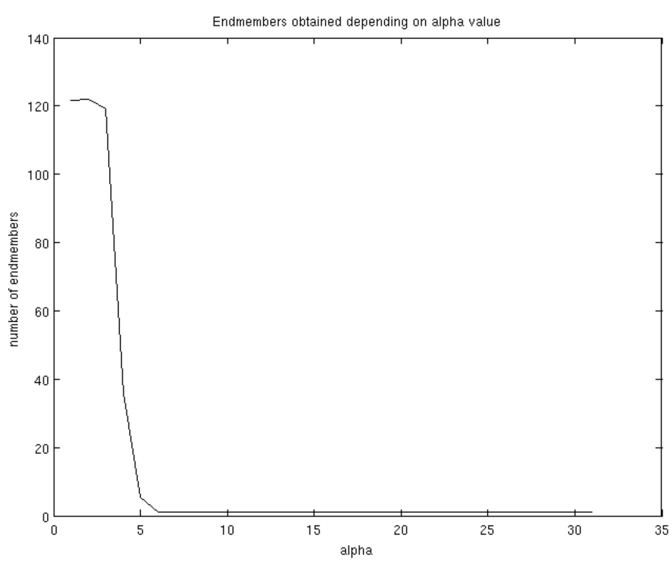

Fig. 5. Number of endmembers retrieved by LICA depending on α.

on both databases show that LICA features perform comparable to both linear feature extraction algorithms. This results open a new computational approach to pattern recognition, specially biometric identification problems. However there are some issues on the LICA algorithm: The uncertainty about the amount of endmembers found and therefore the high variance of recognition rates.

Future works will follow these lines:

- Confirm obtained results performing this same experiment over more complex databases like FERET.
- Combine the non-linear algorithm LICA with other well known statistical tools like PCA, LDA, and other state-of-the art face recognition approaches.
- Work on Lattice Theory mathematical foundations in order to apply energy function-like methods to Lattice Computing implementations that may allow more robust endmember induction.
- Test LICA's capabilities of dealing with face recognition well known problems: Illumination, pose, occlusion, etcetera.

References

1. P. N. Bellhumer, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(7):711–720, 1997.
2. Deng Cai, Xiaofei He, and Jiawei Han. Semi-supervised discriminant analysis. *Computer Vision. IEEE 11th International Conference on Computer Vision*, 14:1–7, 2007.
3. Rama Chellappa, Pawan Sinha, and P. Jonathon Phillips. Face recognition by computers and humans. *IEEE Computer*, 43(2):46–55, 2010.
4. M. Graña, D. Chyzhyk, M. García-Sebastián, and C. Hernández. Lattice independent component analysis for functional magnetic resonance imaging. *Information Sciences*, In Press, Corrected Proof:–, 2010.
5. M. Graña, A. Manhaes-Savio, M. García-Sebastián, and E. Fernandez. A lattice computing approach for on-line fmri analysis. *Image and Vision Computing*, 28(7):1155–1161, 2010.
6. M. Graña, I. Villaverde, J.O. Maldonado, and C. Hernandez. Two lattice computing approaches for the unsupervised segmentation of hyperspectral images. *Neurocomputing*, 72(10-12):2111–2120, 2009.
7. Xiaofei He and Partha Niyogi. Locality preserving projections. In *Proceedings of the Conference on Advances in Nerual Information Processing Systems*, 2003.
8. Guillaume Heusch and Sebastien Marcel. A novel statistical generative model dedicated to face recognition. *Image and Vision Computing*, 28(1):101 – 110, 2010.
9. Sarvani Kare, Ashok Samal, and David Marx. Using bidimensional regression to assess face similarity. *Machine Vision and Applications*, 21(3):261–274, 2008.
10. J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. Face recognition using kernel linear discriminant algorithms. *IEEE Trans. on Neural Networks*, 14(1):117–126, January 2003.
11. J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, and S.Z. Li. Ensemble-based discriminant learning with boosting for face recognition. *IEEE Transactions on Neural Networks*, 17(1):166–178, January 2006.

12. A.V. Nefian. Embedded bayesian networks for face recognition. In *Proc. of the IEEE International Conference on Multimedia and Expo*, volume 2, pages 133–136, Lusanne, Switzerland, August 2002.
13. Lishan Qiao, Songcan Chen, and Xiaoyang Tan. Sparsity preserving discriminant analysis for single training image face recognition. *Pattern Recognition Letters*, 31(5):422 – 429, 2010.
14. Chuan-Xian Ren and Dao-Qing Dai. Incremental learning of bidirectional principal components for face recognition. *Pattern Recognition*, 43(1):318 – 330, 2010.
15. Gerhard X. Ritter and Gonzalo Urcid. A lattice matrix method for hyperspectral image unmixing. *Information Sciences*, In Press, Corrected Proof:–, 2010.
16. G.X. Ritter, P. Sussner, and J.L. Diaz de Leon. Morphological associative memories. *Neural Networks, IEEE Transactions on*, 9(2):281 –293, mar. 1998.
17. G.X. Ritter, G Urcid, and Schmalz M.S. Autonomous single-pass endmember approximation using lattice auto-associative memories. *Neurocomputing*, 72(10-12):2101–2110, 2009.
18. F.S. Samaria and A.C. Harter. Parameterisation of a stochastic model for human face identification. In *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*, pages 138 –142, December 1994.
19. M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neurosicence*, 3(1):71–86, 1991.
20. W.S. Yambor. *Analysis of PCA-based and Fisher Discriminant-Based Image Recognition Algorithms*. Technical report cs-00-103, Computer Science Department, Colorado State University, July 2000.
21. Dake Zhou and Zhenmin Tang. Kernel-based improved discriminant analysis and its application to face recognition. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, 14(2):103–111, 2009.
22. S. Zhou and R. Chellappa. Multiple-exemplar discriminant analysis for face recognition. In *Proc. of the 17th International Conference on Pattern Recognition, ICPR'04*, pages 191–194, Cambridge, UK, August 2004.