Pattern Recognition 45 (2012) 4438–4450

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

The small sample size problem of ICA: A comparative study and analysis

Weihong Deng [a,*], Yebin Liu [b], Jiani Hu [a], Jun Guo [a]

[a] School of Information and Telecommunications Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, People's Republic of China
[b] Department of Automation, Tsinghua University, Beijing 100084, People's Republic of China

ARTICLE INFO

Article history:
Received 21 October 2011
Received in revised form
4 June 2012
Accepted 16 June 2012
Available online 26 June 2012

Keywords:
Projection pursuit
Independent component analysis
Whitened principal component analysis
Locality pursuit
Face recognition

ABSTRACT

On the small sample size problems such as appearance-based recognition, empirical studies have shown that ICA projections have trivial effect on improving the recognition performance over whitened PCA. However, what causes the ineffectiveness of ICA is still an open question. In this study, we find out that this small sample size problem of ICA is caused by a special distributional phenomenon of the high-dimensional whitened data: all data points are similarly distant, and nearly perpendicular to each other. In this situation, ICA algorithms tend to extract the independent features simply by the projections that isolate single or very few samples apart and congregate all other samples around the origin, without any concern on the clustering structure. Our comparative study further shows that the ICA projections usually produce misleading features, whose generalization ability is generally worse than those derived by random projections. Thus, further selection of the ICA features is possibly meaningless. To address the difficulty in pursuing low-dimensional features, we introduce a locality pursuit approach which applies the locality preserving projections in the high-dimensional whitened space. Experimental results show that the locality pursuit performs better than ICA and other conventional approaches, such as Eigenfaces, Laplacianfaces, and Fisherfaces.

1. Introduction

In pattern classification, feature extraction is defined as a mapping from a typically high-dimensional data space to space of reduced dimension while preserving the class separability [1]. PCA and ICA are the two most widely used unsupervised feature extraction techniques. PCA minimizes second-order dependency of the input data to find the basis along which the data (when projected onto them) have maximal variance. ICA minimizes both second-order and higher-order dependencies to find the basis along which the data are statistically independent. PCA is optimal for gaussian signals only, because it neglects the extra information contained in the higher-order statistics. In contrast, ICA uses this higher-order statistical information and is good at describing nonGaussian data.

In the area of appearance-based face recognition, Bartlett et al. claimed that a lot of important information might be contained in the high-order relationships among features (pixels) [2], and thus ICA was commonly considered as a more powerful tool than PCA. Several studies have been conducted for face recognition using ICA algorithm, namely independent Gabor feature method [3], for enhanced ICA by selecting PCA dimension [4]. ICA were also combined with LDA for face recognition [5] and gender classification [6]. Although most empirical studies [2–4] have claimed ICA is better than PCA for feature extraction in the high-dimensional classification system, some studies [7,8] reported contradictory results.

In high-dimensional applications, the ICA pipeline actually contains PCA process (for dimension reduction), whitening process (for scale normalization), and pure ICA process.[1] Yang et al. used the "PCA+whitening" (whitened PCA) as the baseline to revaluate the ICA-based face recognition systems, and the experimental results showed that the performance of ICA is nearly identical to that of whitened PCA [13]. In other words, pure ICA projection has trivial effect on the recognition performance. Based on similar experimental results, Vicente et al. [14] further pointed out that, if all the ICA projections are used, the feature vector derived by ICA is just a rotation of the whitened data, which is meaningless for classification. Therefore, the contradictory results between PCA and ICA can be explained by the effect of whitening process on different data sets. On many data sets, whitening process is effective to improve PCA-based recognition performance [15–18]. The studies used these data sets would report ICA is better than PCA. In some other cases, however, whitening process would lead to overfitting, hence it is not surprising that ICA is inferior.

* Corresponding author. Tel.: +86 10 62283059; fax: +86 10 62285019.
E-mail addresses: whdeng@bupt.edu.cn, whdeng@it.usyd.edu.au (W. Deng).

[1] Throughout this paper, we use FastICA as a representative of various ICA algorithms. Previous studies have shown that the performance difference between FastICA and other ICA implementations, such as Informax [9] and Common's algorithm [10], is not significant [11–13].

0031-3203/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2012.06.010

The equivalence between ICA and whitened PCA is based on the special condition that all the extracted ICA projections are used for classification. In general, ICA is commonly considered a variant of projection pursuit, and a subset of ICA projections can be selected for classification. The usefulness of ICA projections for pattern recognition is often illustrated by some toy samples like Fig. 1, where the projection direction with maximum non Gaussianity clearly highlight the clustered structure of the data. The projection on the first principal component, on the other hand, fails to show this structure. Hence, it is widely believed that selecting a subset of the ICA projections for feature extraction can significantly improve the classification performance [14]. However, the low-dimensional examples like Fig. 1 are not sufficient to verify the effectiveness of ICA on the high-dimensional applications such as the appearance-based recognition, because the high-dimensional data have fundamentally different distribution property from the low dimensional ones.

In this paper, we reveal a *small sample size problem of ICA*: For the high-dimensional data sets, ICA algorithms tend to extract the independent features simply by the projections that isolate single or very few samples apart and congregate all other samples around the origin, without any concern on the clustering structure. To address the difficulty in pursuing low-dimensional features, we introduce two alternative approaches: random pursuit and locality pursuit (LP). Further, we perform a comparative study on ICA, random pursuit, locality pursuit, as well as other state-of-the art dimension reduction methods. Specifically, the contributions of this paper are as follows:

1. We justify that under small sample size condition, *the pairwise distances of the high-dimensional whitened data points are identical*. In other words, the whitening procedure can strictly uniform the pairwise distance between samples, regardless of the intrinsic distribution of the data. As illustrated in Fig. 2, there are three images in the high-dimensional space, and the whitening process maps them onto the vertexes of a equilateral triangle. This finding of the *equaldistant whitened space* unveils a special property of the whitening procedure in the small sample size situation, which brings a new understanding of whitening process beyond the data scaling.
2. We show that the failure of ICA roots from the similarly distant data distribution in the high-dimensional whitened space, where the non-Gaussianity measures of ICA tends to derive the projection directions that isolate a very small number of (even one) data point apart and collapse the others near the origin. To convictively evaluate the applicability of ICA projections, we apply a random projections based algorithm as the baseline, and empirically find that *the ICA projections are less discriminative than the random projections in the high-dimensional whitened space*, which indicates that the ICA model is somehow misleading (worse than random) for high-dimensional classification problems.

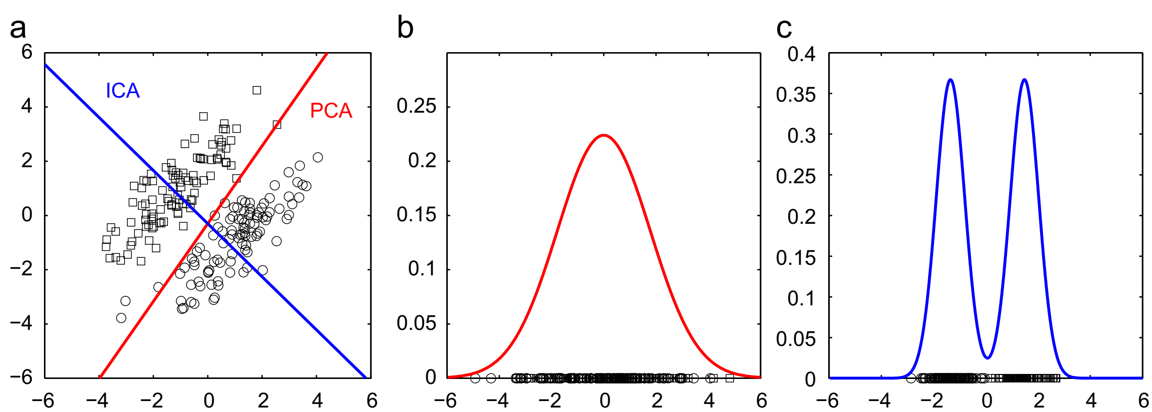

Fig. 1. An illustration of projection pursuit and the “interestingness” of non-Gaussian projections. The data in this figure is clearly divided into two clusters. However, the principal component, i.e. the direction of maximum variance, provides no separation between the clusters. In contrast, the strongly non-Gaussian projection pursuit direction provides optimal separation of the clusters. (a) Projection directions of PCA and ICA. (b) PCA projections. (c) ICA projections.

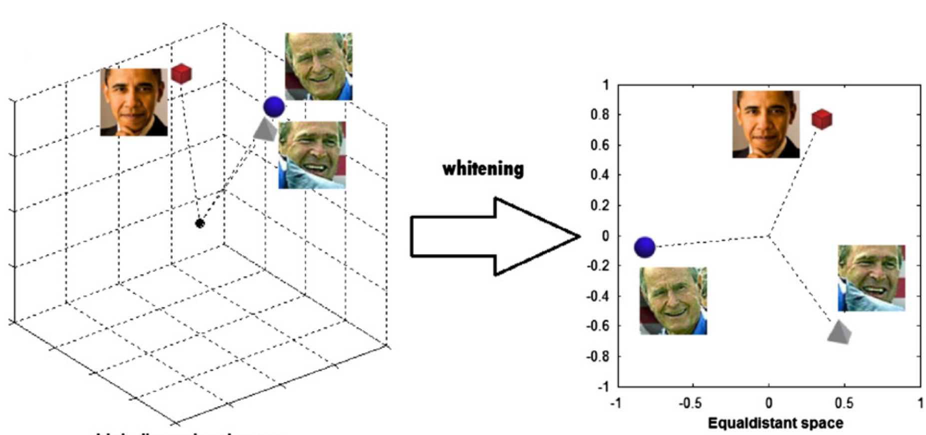

Fig. 2. Illustration of the high-dimensional whitening. The whitening has an unique effects on the high-dimensional data: uniforming the pairwise distance.