Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 40 (2007) 2475–2484
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Boosted manifold principal angles for image set-based recognition

Tae-Kyun Kim*, Ognjen Arandjelović, Roberto Cipolla

Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK

Received 29 June 2006; received in revised form 31 October 2006; accepted 29 December 2006

Abstract

In this paper we address the problem of classifying vector *sets*. We motivate and introduce a novel method based on comparisons between corresponding vector subspaces. In particular, there are two main areas of novelty: (i) we extend the concept of principal angles between linear subspaces to manifolds with arbitrary nonlinearities; (ii) it is demonstrated how boosting can be used for application-optimal principal angle fusion. The strengths of the proposed method are empirically demonstrated on the task of automatic face recognition (AFR), in which it is shown to outperform state-of-the-art methods in the literature.

Keywords: Face recognition; Manifolds; Image set; Principal angle; Canonical correlation analysis; Boosting; Nonlinear subspace; Illumination; Pose; Robustness; Invariance

1. Introduction

Many computer vision tasks can be cast as learning problems over vector *sets*. In object recognition, for example, a set of vectors may represent a variation in an object's appearance—be it due to camera pose changes, non-rigid deformations or variation in illumination conditions. The objective of this work is to classify a novel set of vectors to one of the training classes, each also represented by a vector set. In this paper, learning concepts will be illustrated on sets of face appearance images using the AFR paradigm, although the reader should note that no domain-specific information is actually used.

1.1. Previous work

Most of the previous work on matching vector or image sets exploits their semantics to a certain degree, for example by modelling temporal coherence between consecutive vectors i.e. by matching sequences. By their nature, these methods are of little relevance to the work presented in this paper, so we do not address them here. Broadly speaking, in the recent literature we recognize two groups of approaches to learning over sets of vectors: statistical and principal-angle based.

1.1.1. Statistical methods

Statistical learning approaches rely on the assumption that vectors $\mathbf{x}$ of the ith class are independently and identically (i.i.d.) drawn samples from $p^{(i)}(\mathbf{x})$. The problem of set matching then becomes that of estimating each underlying probability density and comparing two such estimates. In the work of Shakhnarovich et al. [1], densities $p^{(i)}(\mathbf{x})$ are modelled as multivariate Gaussians, estimated with probabilistic principal component analysis (PCA) [2] and compared using the Kullback–Leibler (KL) divergence [3]. Arandjelović et al. criticized this approach for its insufficiently expressive modelling and proposed a kernel-based method to implicitly model nonlinear, but intrinsically low-dimensional manifolds of faces [4]. In this work, the authors also argue against the use of KL divergence due to its asymmetry and demonstrate a superior performance of the resistor–average distance [5] on the task of AFR under mildly varying imaging conditions. In Ref. [6], a Gaussian mixture model (GMM) is proposed for high-dimensional density estimation. The advantage of this approach over the previously mentioned kernel method lies in its more principled modelling of densities confined to nonlinear manifolds; however this benefit comes at the cost

* Corresponding author. Tel.: +44 791 3925831.
E-mail address: tkk22@eng.cam.ac.uk (T.-K. Kim).

0031-3203/$30.00 © 2007 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2006.12.030

of increased difficulty of divergence computation, performed using a Monte–Carlo algorithm.

1.1.2. Principal angle-based methods

Principal angles are minimal angles between vectors of two subspaces (see Section 2). Since the concept of principal angles was first introduced by Hotelling in Ref. [7], it has been applied in various fields [8–10]. Of most relevance to the work addressed in this paper is the mutual subspace method (MSM) of Yamaguchi et al. [11]. In MSM the sum of cosines of the first (i.e. smallest) few principal angles[1] is used as a similarity measure between linear subspaces used to compactly characterize vector sets. MSM has been successfully used for face recognition [11] and ship identification [12] (for evaluation results also see Refs. [4,6]). In the related works [32,13], vector sets are projected to the linear subspace that attempts to maximize the separation (in terms of principal angles) between vector spaces corresponding to different classes, under the assumption of their linearity.

MSM-based methods have two major shortcomings: the limited capability of modelling nonlinear pattern variations and the *ad hoc* fusion of information contained in different principal angles. The assumption of linearity of modelled vector subspaces is important, both because it means that MSM is incapable of differentiating between two nonlinear manifolds embedded in the same linear space and because of the sensitivity of such estimate to particular data variation [4]. In Ref. [14] Wolf and Shashua show how principal angles between nonlinear subspaces can be computed using the "kernel trick" [15]. However, the reported evaluation was performed on a database of a rather small size, making it difficult to judge the performance of their method. Additionally, as in all kernel approaches, finding the optimal kernel function is a difficult problem.

An attractive feature of MSM-based methods is their computational efficiency: principal angles between linear subspaces can be computed rapidly [16], while the estimation of linear subspaces can be performed in an incremental manner [17–20].

1.1.3. Densities vs. subspaces

As a conclusion to this section, we would like to briefly discuss the advantages and disadvantages of the two learning approaches: one which learns densities confined to low-dimensional subspaces and the other which learns the subspaces themselves. In many computer vision applications, due to different data acquisition conditions, the frequency of occurrence of a particular pattern can vary arbitrarily between the training stage and a novel input to the system.[2] In this case, subspace learning techniques are more applicable as they effectively place a uniform prior over a space of possible pattern variation. On the other hand, if there is a reason to believe that training and novel data share some statistical properties, density-based methods may produce better results. In AFR work of Arandjelović et al. [6], for example, the authors note that anatomical constraints and the constraints of the imaging setup make certain head poses more likely than others, therefore opting for a statistical approach to recognition. The point to take is that neither of the two approaches is inherently the right one, but that the choice between the two is dictated by a particular problem.

2. Boosted manifold principal angles (BoMPA)

In this work (the earlier conference version appeared in Ref. [33]), we are interested in discriminating between abstract classes represented as vector sets without any knowledge of what the data represents. Before tackling this problem, it is important to recognize the difficulties of comparing vector sets common to its different semantic instances:

- *Expressiveness*: Pattern changes across and within modelled vector sets often exhibit significant nonlinearities. Seeing that differences within a class can oftentimes be greater than between classes (in Euclidean distance sense), it is important to use a model flexible enough to capture this complex variation, see Fig. 1 for an example. In Section 2.3 we achieve this by moving away from the typically used parametric models and formulate a method that uses canonical correlations and Gaussian mixtures matching.
- *Graceful degradation*: The exact vectors used as an input (either as training or test) to a practical system can be expected to vary from time to time, depending on the exact data acquisition protocol employed. In particular, sometimes more and sometimes less data is available. In the context of face recognition, for example, this may be because the user has not assumed certain poses or because face detection has failed. Graceful degradation refers to slow decay in performance of a learning algorithm as less and less data is available. Our canonical angles-based framework is already exhibiting this property in that only the most similar and discriminating regions of two subspaces are actually compared (see Sections 2.1 and 2.2). Further robustness is achieved by our extension of the similarity function to nonlinear manifolds in Section 2.3 by discarding all but the most reliable matching linear patches.
- *Robustness to noise*: Noise is very much an inherent problem in any practical application. In computer vision, for example, vector patterns considered may represent appearance images—these are affected by noise sources such as quantum, quantization or due to spatial discretization. Our assumption of intrinsically low-dimensional pattern variations within a set, corrupted by isotropic Gaussian noise, are captured well using probabilistic PCA in Section 2.3.
- *Numerical stability and efficiency*: Closely related to the previously mentioned issue of noise in data are numerical issues pertaining to the implementation of a particular algorithm. It is an imperative for a practical algorithm to be numerically stable and, often, be time efficient. These issues are discussed in Sections 2.3 and 3.

[1] In statistics, the cosines of canonical angles are termed canonical correlations.

[2] The term "arbitrarily" should be taken in practical terms i.e. given the parameters which one can realistically expect to model, control or affect.