Pattern Recognition Letters 68 (2015) 190–196

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Grassmann manifold for nearest points image set classification☆

Hengliang Tan[a], Zhengming Ma[a,*], Sumin Zhang[a,b], Zengrong Zhan[a,c], Beibei Zhang[a], Chenggong Zhang[a]

[a] School of Information Science and Technology, Sun Yat-sen University, Guangzhou 510006, PR, China
[b] School of Mathematics and Information, South China Agricultural University, Guangzhou 510642, PR, China
[c] School of Information Engineering, Guangzhou Panyu Polytechnic, Guangzhou 511483, PR, China

ARTICLE INFO

Article history:
Received 17 February 2015
Available online 9 October 2015

Keywords:
Image set classification
Grassmann manifold
Reproducing kernel Hilbert space
Affine hull model
Nearest points

ABSTRACT

Image set classification has attracted increasing attention in recent years. How to effectively represent image sets is one key issue of set based classification. Subspaces form non-Euclidean Riemannian manifolds known as Grassmann manifolds, which allows an image set to be conveniently represented as a point on a Grassmann manifold is widely used in many visual classification tasks. Another issue is how to measure the distance/similarity between sets. Modeling image sets as hulls, and then finding distance of nearest points between sets as the set-to-set distance is a popular solution recently. In this paper, we propose a novel approach by exploiting the Projection kernel that explicitly maps the subspaces from the Grassmann manifold to a Reproducing Kernel Hilbert Space (RKHS) where the Euclidean geometry applies. And then, by modeling the points on RKHS as affine hulls, the Euclidean distance between the nearest points of two hulls can be used for classification. In order to obtain enough points for building the Grassmann affine hulls, we also develop a subspaces constructing method extended by K-means. Experiments are conducted on six datasets. Our proposed method achieves the best classification results on two multi-view object categorization datasets and one extreme illumination variation face recognition dataset.

1. Introduction

With the rapid developments of internet and video surveillance, collecting image sets from video sequences or photo albums is easier, and has many resources. Image set classification in computer vision and pattern recognition has received growing attention [1,4,12,13,17,28–30,32,34,37,38] in recent years. Meanwhile, set based classification has shown superior performance than the single instance based classification approaches in the same conditions, since it provides much more within-set information. The wide range applications of image set classification include object categorization, face recognition, action [9,25] recognition, gesture recognition [2] and texture categorization [10,15], etc.. In this paper we specifically focus on the object categorization and face recognition tasks. Many approaches [1,7,30] have been successfully applied on image set based object categorization and face recognition. However, due to the low resolution and large variations of viewpoint and illumination, these tasks are still challenging for classification based on image sets.

The key issues of image set based classification are how to represent image sets and how to measure the distance/similarity between sets [13]. Generally speaking, addressing the problem of image set classification can be classified into two types, parametric and nonparametric methods. Parametric methods [20,24] model each image set as a parametric distribution and measure the similarity between two distributions. The nonparametric methods have been shown to be superior to parametric methods when the data distributions do not satisfy the models estimated by the parametric methods. A serial of nonparametric methods have been proposed for set based classification. The linear subspace based methods [34,4] model image sets as linear subspaces, and then exploit the principal angles to measure the similarity of two linear subspaces. Kim et al. [17] finds the most discriminative canonical correlation between sets through a linear transformation. Since the visual features and models often lie on non-Euclidean spaces, manifold methods are also very popular in image set classification. The method in [28] represents image set as multiple local linear subspaces and treats them as points on manifold, and then defines manifold-to-manifold distance for sets matching. Manifold Discriminant Analysis (MDA) [29] is proposed to learn an embedding space by maximizing manifold margin. Sparse Approximated Nearest Subspaces (SANS) [2] extracts local linear subspaces from gallery image sets via sparse representation, and then adaptively finds the corresponding closest subspace from the

☆ This paper has been recommended for acceptance by Eckart Michaelsen.

* Corresponding author. Tel.: +86 13925128057.
E-mail addresses: tanhengl@mail2.sysu.edu.cn (H. Tan), zhengmingma2015@gmail.com, issmzm@mail.sysu.edu.cn (Z. Ma).

http://dx.doi.org/10.1016/j.patrec.2015.09.008
0167-8655/

samples of a probe image set by joint sparse representation; the distance between subspaces over Grassmann manifolds is defined by the Frobenius norm distance. Recent studies have shown that better performance can be achieved when the geometry of Riemannian space is explicitly considered [7,9,15,30,26]. In [7] and [31], Grassman kernels are used to map the subspaces to Reproducing Kernel Hilbert Spaces (RKHS) where Euclidean geometry applies, classification is achieved with the Kernel Linear Discriminant Analysis (KLDA). A similar method [8] uses a graph embedding discriminant analysis to generalize the method of [7]. Symmetric Positive Definite (SPD) matrices which lie on Riemannian manifolds [19] are also used for image set classification. In [30] and [15], covariance matrices are employed to represent image sets, and then they are mapped to a high dimensional Hilbert space, so that, such as Support Vector Machine (SVM), Kernel Principal Component Analysis (KPCA) and KLDA can be adopted for classification. However, the SPD matrices based methods often suffer from the curse of dimensionality and computational complexity [38].

The nonparametric methods mentioned above are more attentive on the representation of image sets, another serial of nonparametric approaches which concentrate more on how to measure the distance between sets are the hull based methods. Affine/Convex Hull based Image Set Distance (AHISD/CHISD) [1] represents images as points in a linear or affine feature space, and then computes the distance of convex geometric region spanned by its feature points. Hu et al. [13] propose Sparse Approximated Nearest Points (SANP), which sparse representation is applied to regularize the affine hull model. The Regularized Nearest Points (RNP) [35] is proposed to reduce the model complexity of SANP. After that, Wu et al. [33] and Zhu et al. [38] propose methods by employing the collaborative representation technique to utilize the discrimination information between gallery sets, which further improve the performances. These approaches actually aim to find the synthetic nearest points between image sets. However, sometimes the synthetic nearest points between two sets estimated by the affine combination of the existing samples are not consistently meaningful [23]. Furthermore, complex appearance variations caused by multiple views and extreme illumination often generate nonlinearity [22]. Although the kernel trick with Gaussian kernel has been exploited to handle the nonlinearity, such as the kernel version of AHISD [1], kernel SANP [14] and the Kernelized Convex Hull based Image Set Collaborative Representation and Classification (KCH-ISCRC) [38], but as we can see in the experiments of these literature studies, the improvements are limited. Very recently, a novel deep learning framework for image set classification has been proposed [11,12]. However, deep learning often casts great demand to the computation. And their work is not robust to noisy image data, outliers and diverse within-set data variations.

In order to solve the problems mentioned above, as the visual data often lies on the non-Euclidean space and inspired by the kernel AHISD. In this work, we proposed a novel approach called Grassmann Nearest Points (GNP) for a solution. We first develop the K-means clustering method to cluster each image set into multiple overlapped local patches, each patch can be modeled as a linear subspace which is a point on a Grassmann manifold. And then, these subspaces are mapped to an RKHS by Grassmann kernel, where many algorithms on Euclidean spaces can be directly generalized. After that, we find the distance between nearest points of two sets on the mapped space by the affine hull model. These can be technically achieved by the kernel trick with Grassmann kernel. The proposed GNP is to the best of our knowledge the first one that uses the kernel affine hull model with Grassmann kernel to deal with image set classification. Experiments are conducted on two multi-view object categorization datasets and four face recognition datasets. The experimental results show that the proposed GNP gives notable performances on solving the multi-view and extreme illumination variation problems comparing to several state-of-the-art methods.

The rest of this paper is organized as follows. In Section 2, we briefly introduce the Grassmann manifold and how to construct subspaces in one image set. Section 3 reviews the affine hull model and presents our proposed Grassmann nearest points method. Experimental results are presented in Section 4. Section 5 concludes this paper.

2. Modeling image sets on Grassmann manifold

This section we introduce the Grassmann manifold and Grassmann kernels. Following that, our subspaces construction method is presented.

2.1. Grassmann manifold

Manifold analysis has been extensively studied with success in wide range of research. Manifold can be considered as topological space that is locally similar to Euclidean space and has a globally defined differential structure. Meanwhile, manifold can be embedded in a higher dimensional reproducing kernel Hilbert space, where many Euclidean algorithms can be generalized [15]. Grassmann manifold is a Riemannian manifold that embedded in a higher dimensional RKHS, and it's formed by subspaces. In this paper, we specially focus on this type of manifolds.

Given a set of m-dimensional linear subspaces of $\mathbb{R}^D$ and the Grassmann manifold $\mathcal{G}_{D,m}$. An element of $\mathcal{G}_{D,m}$ can be represented by a linear subspace that spanned by an orthonormal matrix $\mathbf{Y}_i$ with the size of $D \times m$, which is span($\mathbf{Y}_i$). Then, the Riemannian distance between two subspaces on Grassmann manifold can be defined, such as the Projection metric and Binet-Cauchy metric [7] which are based on the principal angles. Due to the specific geometric properties of Grassmann space, Grassmann manifold can be mapped to an RKHS by using Grassmann kernels which obey Mercer's theorem. Let the kernel function $k : \mathcal{G} \times \mathcal{G} \to \mathbb{R}$ be a symmetric real-valued function, where k is a Grassmann kernel if and only if it satisfies the positive definiteness and being well-defined [7]. Various Grassmann kernels have been successfully applied for computer vision tasks [7,8]. In this work, we are interested in the Projection kernel, which is generalized by the Projection metric. The Projection kernel has achieved consistent and promising results in literature [7–9]. The Projection kernel is given by:

$$k_P(\mathbf{Y}_1, \mathbf{Y}_2) = \left\| \mathbf{Y}_1^T \mathbf{Y}_2 \right\|_F^2 \tag{1}$$

Where $\| . \|_F$ denotes the Frobenius norm, $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are two orthonormal matrices with the same size $D \times m$.

2.2. Construction of subspaces on an image set

In many challenging visual classification scenes, there is only one image set in each class for gallery or training set. However, for many classification tasks, such as the discriminant analysis methods, they need at least two training objects (or sets) for discriminative learning. One convenient way to solve this problem is by dividing the one image set into several subsets. As mentioned previously, Grassmann manifolds take linear subspaces as points on non-Euclidean spaces. Since the linear subspaces are able to accommodate the effects of complex data variations [9,30], modeling image sets as linear subspaces have been proven to be beneficial for many visual classification tasks.

In this work, different from the method in [7] which models each image set as one linear subspace, we model an image set as multiple subspaces by a simple set dividing method which extended by K-means clustering. Modeling an image set as multiple subspaces has been previously proposed, such as [28,29]. They use two similar types of Maximal Linear Patch (MLP) algorithms respectively to cluster an image set into several non-overlapped local patches, and then extract