Image and Vision Computing xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Image and Vision Computing

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

Joint regularized nearest points for image set based face recognition☆

Meng Yang [a,b,]*, Xing Wang [b], Weiyang Liu [c], Linlin Shen [b]

[a] School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China
[b] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
[c] School of Electronic & Computer Engineering, Peking University, China

ARTICLE INFO

Article history:
Received 18 September 2015
Received in revised form 30 June 2016
Accepted 16 July 2016
Available online xxxx

Keywords:
Face recognition
Image set
Joint regularized nearest points
Sparse representation

ABSTRACT

Face recognition based on image set has attracted much attention due to its promising performance to overcome various variations. Recently, classifiers of regularized nearest points, including sparse approximated nearest points (SANP), regularized nearest points (RNP) and collaborative regularized nearest points (CRNP), have achieved state-of-the-art performance for image set based face recognition. From a query set and a single-class gallery set, SANP and RNP both generate a pair of nearest points, between which the distance is regarded as the between-set distance. However, the computing of nearest points for each single-class gallery set in SANP and RNP ignores collaboration and competition with other classes, which may cause a wrong-class gallery set to have a small between-set distance. CRNP used collaborative representation to overcome this shortcoming but it doesn't explicitly minimize the between-set distance. In order to solve these issues and fully exploit the advantages of nearest points based approaches, in this paper a novel joint regularized nearest points (JRNP) is proposed for face recognition based on image sets. In JRNP, the nearest point in the query set is generated by considering the entire gallery set of all classes; at the same time, JRNP explicitly minimizes the between-set distance of the query set and a single-class gallery set. Furthermore, we proposed algorithms of greedy JRNP and adaptive JRNP to solve the presented model, and the classification is then based on the joint distance between the regularized nearest points in image sets. Extensive experiments were conducted on benchmark databases (e.g., Honda/UCSD, CMU Mobo, You Tube Celebrities databases, and the large-scale You Tube Face datasets). The experimental results clearly show that our JRNP leads the performance in face recognition based on image sets.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Recognizing the objects of interest (e.g., human faces) is one of the most important and fundamental problems in the communities of computer vision and pattern recognition. Although face recognition (FR) has been extensively studied in the past several decades, traditional face recognition works usually assume there is only a single query face image, from which a human identity is recognized [1–3]. Although there are multiple training images in the gallery set per subject, it is still a big challenge to correctly recognize a person's identity based on only a single query face image with various variations (e.g., lighting, expression, pose, disguise changes) captured in less-controlled/uncontrolled environments. As shown in Fig. 1, we can observe that the single query face image is quite different from the images in the gallery set, which although has already contained many samples with different variations.

With the wide installation of video cameras and the developments of large-capacity-storage media, it becomes very convenient to collect multiple images from video sequences or photo albums for a known subject and store these images as the gallery and query image sets. Multiple face images in the query and gallery set for each subject incorporate more different intra-class appearance variations and provide richer discrimination information than a single query face image for face recognition. Compared to the traditional face recognition with a single query face image, face recognition based on image sets can achieve more satisfactory performance in practical face recognition applications and is a more promising framework of face recognition.

Face recognition based on image sets has been attracting much attention of researchers for the past decades. The image sets are either the consecutive video sequences with temporal information, or unordered pictures (e.g., photo album images collected from web at different times). Compared to video-based face recognition [4–10], face recognition based on general image sets, in which the temporal information is not available, has wider applications (e.g., both video based face recognition and face recognition based arbitrarily collected images). In this paper we mainly focus on the face recognition problem based on

☆ This paper has been recommended for acceptance by Maja Pantic.

* Corresponding author.
E-mail address: yang.meng@szu.edu.cn (M. Yang).

http://dx.doi.org/10.1016/j.imavis.2016.07.008
0262-8856/© 2016 Elsevier B.V. All rights reserved.

ARTICLE IN PRESS

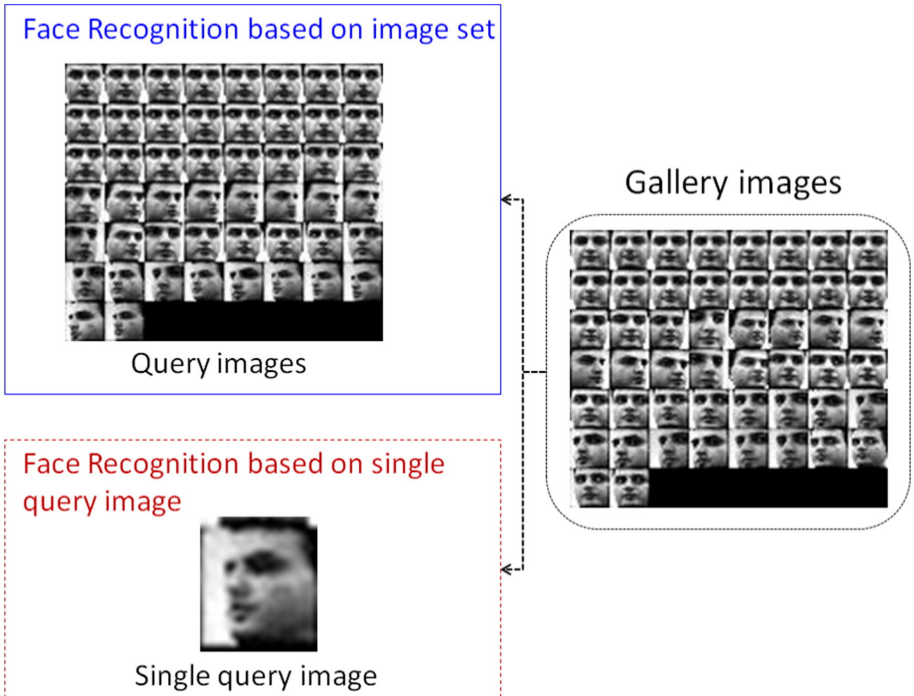

Fig. 1. Face recognition based on image set and traditional face recognition based on a single query image.

general image sets. Numerous approaches have been proposed for this kind of image-set based recognition problem.

One major category of face recognition based on image set is parametric model based approaches [10–12]. These parametric model based approaches [10–12] firstly represent each image set in both the gallery set and the query image set by some parametric distribution with its parameters estimated from the image set data itself, and then calculate the between-set distance by measuring the similarity between these two distributions (e.g., in terms of Kullback–Leibler divergence [13]). However, the parametric methods need to solve a difficult parameter estimation problem and require strong statistical correlations between the gallery and query sets, which may not exist in practice. To overcome the shortcomings of parameter model based approaches, recently Lu et al. [14] directly extracted the multiple order statistics features from each image set and developed a multi-kernel metric learning method to combine different order information.

In order to avoid the drawbacks of model-based methods, non-parametric model-free based approaches were proposed by representing an image set as a convex/affine subspace [15–19], a hierarchical-structure subspaces [20], locally orthogonal subspaces [21], local models via clustering [22], or nonlinear manifolds [23–26]. When there are many samples in a data set, the dense samples can construct a manifold with a complicated surface. Nevertheless, each point of a manifold has a local Euclidean space, which has been widely applied in manifold-based face recognition. In nonlinear manifold methods, the nonlinear manifold of an image set is liberalized locally and divided into a set of local linear subspaces [23,26], of which each can represent a local surface of the manifold. In this model-free face recognition based on image sets, how to measure between-set distance is the key problem. A popular way is to define the between-set distance of two image sets as the distance between two "exemplars" (e.g., the mean of samples) chosen from these two sets. For instance, Cevikalp et al. [15] characterized each image set by an affine/convex hull spanned by its samples, and selected two points (one point in one hull) with the closest approach as the "exemplars". Another way of measuring the between-set distance for non-parametric approach is to compare the structure of the non-parametric model. For instance, Canonical correlation analysis (CCA) [27], which analyzes the principal angles and canonical correlations between linear subspaces, is widely used in the works of [16–19,21–23]. Besides, the natural second-order statistic–covariance matrix was used to represent each image set in [28], and the image-set based classification was formulated as classifying points lying on a Riemannian manifold.

Recently inspired by the success of sparse representation on face recognition [29], Hu et al. [30] proposed a sparse approximated nearest points (SANP) approach for image-set based face recognition. By modeling each image set as an affine hull, Hu et al. selected two points (one point in each hull) with the closest distance as the sparse approximated nearest points (SANP), where the nearest points of SANP were required to be sparsely represented by the original samples. The final between-set distance of SANP is the result of multiplication of the distance between the generated nearest points of SANP and the total dimensions of the affine hulls of two image sets. The goal of multiplication with the dimensions is to eliminate the bias to the larger-size image set (e.g., with more images) and ensures that a small between-set distance is only obtained when the distance between the two nearest points and the dimensions of sets are both small [30]. Although SANP has achieved a good performance, its computational model (e.g., three representation terms and four unknown variables) is a little complex. In order to improve it, Yang et al. [31] proposed a regularized nearest points (RNP) method, which modeled each image set as a regularized affine hull and use the regularized nearest points to measure the similarity of these two image sets. Compared to the conventional affine hull to model an image set, the affine hull in RNP is a regularized one where the representation coefficients are regularized by using l_2-norm (e.g., it requires that the l_2-norm of the coding coefficient vector is smaller than a scalar constant). The purpose is to avoid containing the meaningless points and make the representation stable [31]. Following RNP and collaborative representation based classification [32], Wu et al. [33] find all the regularized nearest points simultaneously in the framework of collaborative representation and propose the collaboratively regularized nearest points (CRNP) method. Similar to [33] to borrow the idea of collaborative representation, Zhu et al. [43] also proposed a regularized hull based image set based collaborative representation

Please cite this article as: M. Yang, et al., Joint regularized nearest points for image set based face recognition, Image Vis. Comput. (2016), http://dx.doi.org/10.1016/j.imavis.2016.07.008

Download English Version:

https://daneshyari.com/en/article/4969032

Download Persian Version:

https://daneshyari.com/article/4969032

Daneshyari.com