Pattern Recognition 48 (2015) 2935–2946

Contents lists available at ScienceDirect

Pattern Recognition

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

Mixed-norm sparse representation for multi view face recognition

Xin Zhang [a,*], Duc-Son Pham [b], Svetha Venkatesh [a], Wanquan Liu [b], Dinh Phung [a]

[a] *Centre for Pattern Recognition and Data Analytics, Deakin University, GEELONG, VIC, Australia*
[b] *Department of Computing, Curtin University of Technology, Perth, WA, Australia*

A R T I C L E I N F O

Article history:
Received 6 November 2013
Received in revised form
19 February 2015
Accepted 25 February 2015
Available online 17 March 2015

Keywords:
Multi-pose face recognition
Sparse representation classification
ADMM
Group sparse representation
Multi-task learning
Joint dynamic sparse representation classification
Unsupervised learning
Convex optimization
Robust face recognition

A B S T R A C T

Face recognition with multiple views is a challenging research problem. Most of the existing works have focused on extracting shared information among multiple views to improve recognition. However, when the pose variation is too large or missing, 'shared information' may not be properly extracted, leading to poor recognition results. In this paper, we propose a novel method for face recognition with multiple view images to overcome the large pose variation and missing pose issue. By introducing a novel mixed norm, the proposed method automatically selects candidates from the gallery to best represent a group of highly correlated face images in a query set to improve classification accuracy. This mixed norm combines the advantages of both sparse representation based classification (SRC) and joint sparse representation based classification (JSRC). A trade off between the ℓ_1-norm from SRC and $\ell_{2,1}-$norm from JSRC is introduced to achieve this goal. Due to this property, the proposed method decreases the influence when a face image is unseen and has large pose variation in the recognition process. And when some face images with a certain degree of unseen pose variation appear, this mixed norm will find an optimal representation for these query images based on the shared information induced from multiple views. Moreover, we also address an open problem in robust sparse representation and classification which is using ℓ_1-norm on the loss function to achieve a robust solution. To solve this formulation, we derive a simple, yet provably convergent algorithm based on the powerful alternative directions method of multipliers (ADMM) framework. We provide extensive comparisons which demonstrate that our method outperforms other state-of-the-arts algorithms on CMU-PIE, Yale B and Multi-PIE databases for multi-view face recognition.

1. Introduction

Face recognition is one of the non-intrusive biometrics. Due to the emerging demand in surveillance and security, it is an important research topic in pattern recognition [1]. According to previous literature surveys [2,3], extensive studies have been done to resolve the face recognition issues, such as pose, illumination, expression, and occlusion [4–6]. However, most of these methods are based on a single input image. They identify a subject by matching a single query face image with all gallery images one by one. In practice, it is common that the query face image is noisy or its pose may be missing in the gallery, thus working with a single face image is likely to be unreliable in real-world applications. On the other hand, multiple views of a same subject can be obtained easily with current technology. For instance, a sequence of face images from a subject with a large degree of pose variations may be observed over a time interval by a surveillance camera or multiple snapshots are captured by video camera networks at the same time from different viewpoints. This will produce a large number of query images for recognition tasks. Since multiple view images are from the same subject under different time or viewpoint, there is likely some shared information across those face images. The existing face recognition techniques have not investigated the inter-correlation among the query images, therefore, exploiting the using of these shared information becomes an important work.

In the face recognition literature, several popular classifiers have been developed. The nearest neighbour (NN) is one of the most common and popular classifiers [7]. The NN classifies the query face image based on its closest neighbour in the gallery set. However, this classifier is sensitive to outliers. The NN classifier is generalized to nearest subspace (NS) [8]. Instead of using a single image to perform classification, NS classifies a face based on the best linear representation in terms of all the gallery images in each class. Since the classification decision is made by all samples, NS is more robust than NN. Sparse representation-based classification (SRC) [6] seeks a balance between these two extreme cases, it represents a query image by adaptively choosing a minimum number of atoms (samples in gallery) from both within each class and across multiple classes. SRC

* Corresponding author.
E-mail addresses: xin.zhang@curtin.edu.au (X. Zhang), ducson.pham@curtin.edu.au (D.-S. Pham), svetha.venkatesh@deakin.edu.au (S. Venkatesh), w.liu@curtin.edu.au (W. Liu), dinh.phung@deakin.edu.au (D. Phung).

http://dx.doi.org/10.1016/j.patcog.2015.02.022
0031-3203/

has been shown more robust and effective than NN and NS on some common face recognition issues, such as occlusion and corruption. Encouraged by the SRC framework, a large number of its extensions have been proposed [9–11] and they have achieved state-of-the-art performance. However, they are limited to single query face image for recognition.

Recently, a growing interest [12–15] in face recognition from an image set has emerged. Rather than using a single query image to perform recognition, multiple face images of the same subject are used as an input. In general, the system identifies a query subject based on a set of input images from known subjects in the gallery. The face images in both gallery and query sets may have large variations in pose, illumination, etc. By using multiple face images of the same subject in the query, the robustness of the recognition system has been improved significantly compared with single-input systems. In [16], an extended volume Local Binary Patterns is introduced to exploit the information among frames. It can achieve a good performance, but it requires sequential images from a video. Another approach to achieve this goal is by measuring the distance between the query set and each class in the gallery set. Inspired by label propagation [17], a graph-based classification for multiple view face recognition has been proposed [12]. It converts the face image set into a similarity graph, and then uses a class-wise graph matching procedure to compare this similarity graph with the graph generated by each class in the gallery. In [13], face images are represented as a feature vector in an affine feature space. They build an affine hull for each image set (query set and each class in the gallery). The geometric distances between the affine hull of query and of each class in the gallery are used to make the classification decision. A multi-class group Lasso is introduced in [18]. Images are represented by Local Binary Patterns [19], then the best suitable features are selected to measure the distance between each pair of sets. These methods treat each set of the gallery face images as a linear subspace, and use subspace distance to identify the query subject from subjects in the gallery independently. Thus, they have two limitations: (1) they cannot exploit information across multiple classes; (2) when there is a large difference between images in the same class, such as large pose variations, these methods can perform poorly.

Since SRC considers both within each class and between multiple class factors, a multiple test samples generalization of SRC is introduced, known as joint sparse representation-based classification (JSRC) [14,20]. This method assumes that the query face images share the same sparsity pattern. The shared information can be exploited by using this assumption. Instead of solving the SRC problem for each query image, JSRC solves a set of query images from the same subject. It adaptively selects a minimum number of atoms from gallery images, these atoms can best represent every query images at same time. However, this assumption will not hold when there are large pose differences in the query images. For example, if a frontal face and a 90° right face exist in the query set at the same time, it is impossible to find an atom in the gallery to represent both of them at a same time accurately. In order to overcome this issue, joint dynamic sparse representation-based classification (JDSRC) was proposed in [15]. The authors in [15] argue that the same sparsity patterns is not necessary at the atom level, these patterns should be at the class level. To capture this model, they introduced a new concept of joint dynamic sparsity. This joint dynamic sparsity brings in flexibility to atom selection of JSRC. When the pose variation is large in the query images, JDSRC does not necessarily select the same atom for all poses as JSRC. Instead, JDSRC selects atoms from the whole class to represent all poses. Nevertheless, when a pose appears in the query but is missing in the gallery, JDSRC will be forced to select a 'similar' atom from the gallery to represent it. This may not lead to a robust solution. In addition, the JDSRC is achieved by an extension of simultaneous orthogonal matching pursuit [20] which is a naive greedy method and may be not convergent. Therefore, a new algorithm is needed to solve this challenging multi-view (multi-pose) problem.

In [21], the authors argue that the robustness of SRC based methods should be achieved by using the ℓ_1-loss function instead of the ℓ_2-loss in SRC. However, it was left as an open question, because solving via standard linear programming techniques is computationally expensive. The sparse representation is inspired from compressed sensing (CS). In the statistical signal processing community, the core CS problem is finding a sparse linear combination of signal atoms from an overcomplete dictionary [22,23]. It was then applied to face recognition in [6]. As solving this CS problem is close to the Lasso in statistics in functional form, extensions to the basic sparse solution have been observed in related areas. A robust Lasso, which explicitly models the corruptions, is proposed and analysed in [24]. Statistically, this is more generic and provably better than the least entropy and error correction alternative discussed in a rejoinder [25] by the authors of SRC against the paper of Shi et al. [21]. However, this is obtained at the cost of an extra regularization parameter. In the related robust CS paper [26], a slightly different loss function, known as Huber's robust loss function is used. However, it requires the estimates of the Huber's parameters, which brings additional computational burden.

In this paper, we propose a novel mixed norm sparse representation classification (MSRC) method for multi-view face recognition. The proposed method has the similar ability to JDSRC, it allows some degree of flexibility in atom selection procedure of JSRC. On one hand, as SRC works with a single query image, it cannot exploit the shared sparsity pattern across query images. Thus, it will ignore the influence of large pose variations in the query images. On the other hand, JSRC struggles with shared information among query images, but it can easily be affected by the pose variations. Therefore, it is natural to strike a balance between them. Our MSRC achieves this goal. It exploits the correlation among the variance face images in the query and it also brings the flexibility to the atom selection to achieve an accurate and sparse representation. Moreover, to achieve more robustness, our MSRC uses the ℓ_1-loss instead of the general ℓ_2-loss. Indeed, the ℓ_1-norm loss function we use in this work, which is also an open question discussed in [21], is also known in the robust statistics literature to be optimal for noise modelled as a Cauchy distribution.

The contributions of this work are as follows: (1) we derive a simple, provably convergent, and computationally efficient algorithm based on the framework of alternative direction method of multipliers (ADMM) [27]; (2) we establish a novel multi-view face recognition in a robust form to exploit the similarities between different images in the query images; (3) we provide extensive experiments to show the advantages of our proposed method against other state-of-the-arts.

The paper is organized as follows. We first briefly review the related works in Section 2. Then we derive our multi-pose face recognition via sparse representation (MSRC) based on the ADMM framework in Section 3. Finally, we provide extensive experiments on CMU-PIE, Yale B and Multi-PIE datasets in Section 4. Section 5 concludes the paper.

The Matlab implementation of proposed method is publicly available at https://sites.google.com/site/dspham/code.

2. Related works

2.1. Face recognition via sparse representation

In sparse representation classification (SRC), given a set of gallery images $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N]$ where each $\mathbf{a}_i \in \mathbb{R}^d$, one seeks a