Discriminant Projection Representation-based Classification for Vision Recognition

Qingxiang Feng and Yicong Zhou*
Computer and Information Science, University of Macau
fengqx1988@gmail.com, yicongzhou@umac.mo, *Corresponding author.

Abstract

Representation-based classification methods such as sparse representation-based classification (SRC) and linear regression classification (LRC) have attracted a lot of attentions. In order to obtain the better representation, a novel method called projection representation-based classification (PRC) is proposed for image recognition in this paper. PRC is based on a new mathematical model. This model denotes that the 'ideal projection' of a sample point x on the hyper-space H may be gained by iteratively computing the projection of x on a line of hyper-space H with the proper strategy. Therefore, PRC is able to iteratively approximate the 'ideal representation' of each subject for classification. Moreover, the discriminant PRC (DPRC) is further proposed, which obtains the discriminant information by maximizing the ratio of the between-class reconstruction error over the within-class reconstruction error. Experimental results on five typical databases show that the proposed PRC and DPRC are effective and outperform other state-of-the-art methods on several vision recognition tasks.

Introduction

Recently, representation-based classifiers have attracted increasing attentions of researchers, which can be roughly divided into two kinds: all-classes-based and single-class-based. In the first kind, the well-known method is sparse representation-based classification (SRC) (Wright et al. 2009). It was developed to use the all-class-model to obtain the L_1-based sparse representation for classification . To improve the computation efficiency, the collaborative representation-based classification (CRC) (Zhang, Yang, and Feng 2011) was proposed to address the L_2 minimum problem. Later, several improved methods of SRC were proposed for image recognition (Yang et al. 2012; Feng and Zhou 2016b; Feng and Zhou 2017), such as manifold constraint transfer (MCT) (Zhang et al. 2015) applies a strategy to produce new data for classification. Different from the all-class-model in SRC, some classifiers use single class to obtain its representation. For example, linear regression-based classification (LRC) (Naseem, Togneri, and Bennamoun 2010) was proposed for face identification, which was based on that samples from a specific object class are known to lie on a linear subspace (Basri and Jacobs 2003; Feng, Zhou, and Lan 2016; Feng and Zhou 2016a). LRC solves the least square errors and obtain the linear projection point as the representation for an independent class-specific models.

The common objective of existing representation-based methods is to find the best representation for classification. However, they have only obtained a roughly approximated representation. For example, In ref.(Wright et al. 2009), we know that the ideal representation of SRC is to solve the L_0 minimum problem. In LRC, the regression projection is obtained by a matrix's pseudo-inverse. Therefore, we know that they only obtain the approximated representation. In order to find a better representation of an image, this paper proposes a projection representation-based classification (PRC) for image recognition. To approximate the 'ideal representation' of a sample point, PRC utilizes a new mathematical model to iteratively compute the projection point of the test sample towards a line linking a paired of specific points. This mathematical model has been proved by a theorem. According to the theorem, we know that the generated projection will be almost equal to the 'ideal representation' after sufficient iterations. Moreover, the discriminant PRC (DPRC) is further proposed, which obtains the discriminant information by maximizing the ratio of the between-class reconstruction error over the within-class re-construction error. The main contributions of this paper are as follows:

⊙ Firstly, we propose a new mathematical model to obtain the projection of a point on a hyper-space, and we prove this model mathematically.

⊙ Secondly, with the mathematical model, we propose the projection representation-based classification (PRC) for image recognition tasks. The generated projection by PRC will be almost equal to the 'ideal representation' after sufficient iterations.

⊙ In order to obtain an effective discriminant subspace for PRC, we propose the discriminant PRC (DPRC). DPRC utilizes the labeled training samples set to constitute a more reliable subspace such that the effective discriminant information can be used for classification.

⊙ Experiments have been carried out on several challenging databases. The results show that the proposed PRC and DPRC outperform several state-of-the-art methods.

Notation Summary

Let $X = \{x_i^c \in R^{q\times 1}, i = 1, 2, \cdots, N_c, c = 1, 2, \cdots, M\}$ denote the prototype data set, where x_i^c is the i^{th} sample of the c^{th} class, M is the number of classes, N_c is the number of samples of the c^{th} class and q is the sample's dimension. The number of all the samples is $L = \sum_{c=1}^{M} N_c$. The prototype data set can be also described as $X = \{x_i \in R^{q\times 1}, i = 1, 2, \cdots, L\}$.

Proposed Math Model

Before introducing the math model, we describe the 'ideal projection' in Definition 1.

Definition 1: *Suppose that there exists a test sample x and a specific class c. If a point in the hyper-space of the c^{th} class is the nearest to test sample x, it is treated as the 'ideal representation' or 'ideal projection' of the test sample x on the hyper-space.*

Math Model

Given a point x and a hyper-space H, the 'ideal projection' of x on the hyper-space H may be gained by iteratively computing the projection point of x on a line of hyper-space H with the proper strategy. It can be describe as

$$p\{x, H\} \approx \underset{k\to+\infty}{repreat}\ p\{x, \overline{x_i^{c,k} x_{i^*}^{c,k}}\} \tag{1}$$

where $p\{*, \bigotimes\}$ denotes the projection of $*$ on $\bigotimes$, $x_i^{c,k}$, $x_{i^*}^{c,k}$ are two points of the hyper-space H. $x_{i^*}^{c,k}$ is the nearest point among all known-distance points of H.

Correctness of the Math Model

Now, we know that the proposed model is quite useful for finding the better representation of each class for classification. Therefore, the correctness of the proposed math model will be an important problem. The Theorem 1 is provided to prove the proposed math model. According to the Theorem 1, a projection point with the minimum distance can be obtained by iteratively computing the projection point of x on hyper-space's a line. Considering Definition 1, we know that the obtained projection point can be treated as the projection of x on the hyper-space. Therefore, the proposed math model is correct.

Theorem 1: *Given a test sample x, and a specific class c with N_c training samples. Suppose the c^{th} class in the first round of projection is $X_c^0 = [x_1^{c,0}\ x_1^{c,0}\ \cdots\ x_{N_c}^{c,0}]$. Select the nearest training sample $x_n^{c,0}$ and another training sample $x_i^{c,0}$, $i \neq n$ to form a line $\overline{x_n^{c,0} x_i^{c,0}}$; Compute the projection point $x_p^{c,0}$ of the test sample x to the line $\overline{x_n^{c,0} x_i^{c,0}}$; and use $x_p^{c,0}$ to replace $x_i^{c,0}$, $i \neq n$ as the new training set of the c^{th} class. If this projection operation was performed in sufficient times, the distance between the test sample and new projection point closely approximates to a fixed constant, which is the smallest distance between the test sample x and the space of the c^{th} class.*

Proof: Because $x_p^{c,0}$ is the projection of the test sample x to the line $\overline{x_n^{c,0} x_i^{c,0}}$, then

$$\|x - x_p^{c,0}\| \le \|x - x_n^{c,0}\|. \tag{2}$$

After the first projection procedure, $x_p^{c,0}$ replaces $x_n^{c,0}$ as the new nearest sample and will be used to form the new line. Following this manner, in the k^{th} projection procedure, we have

$$\|x - x_p^{c,k}\| \le \|x - x_p^{c,k-1}\| \tag{3}$$

Because the distance between the test sample x and the projection point is equal or greater than 0, the projection points satisfy the following conditions.

$$0 \le \|x - x_p^{c,k}\| \le \|x - x_p^{c,k-1}\|,\ k = 1, 2, \cdots, +\infty \tag{4}$$

That is,

$$\lim_{k\to+\infty} \|x - x_p^{c,k}\| = d \tag{5}$$

where d is a constant that is the smallest distance between the test sample x and the subspace of the c^{th} class.

Proposed Math Model vs Linear Regression

For a specific class subspace,the real projection point cannot be computed using the existing math knowledge because the class subspace is a hyper-space. Ref. (Naseem, Togneri, and Bennamoun 2010) proposed LRC to solve the least square errors and obtain the linear projection point. LRC has the good performance. However, LRC obtains the linear projection point by a pseudo-inverse operation such that this point is only a roughly approximated projection point (exist the closer point than the linear projection point), not the ideal projection point according to Definition 1. Therefore, we intend to obtain a better projection point that is the nearest one to the 'ideal projection' point by using the proposed math model.

Proposed PRC

Based on the concept of finding the best representation of each class, this section proposes a new classifier, called projection representation-based classification (PRC). According to the proposed math model, PRC may obtain the approximated projection point by computing the projection point of the test sample to a line linking with a pair of training samples iteratively. The flowchart of PRC is shown in Figure 1.

Projection Representation

To find a point extremely close to the ideal representation of a sample point, PRC iteratively computes the projection point of the test sample on a line. The final result will be treated as the projection representation for classification.

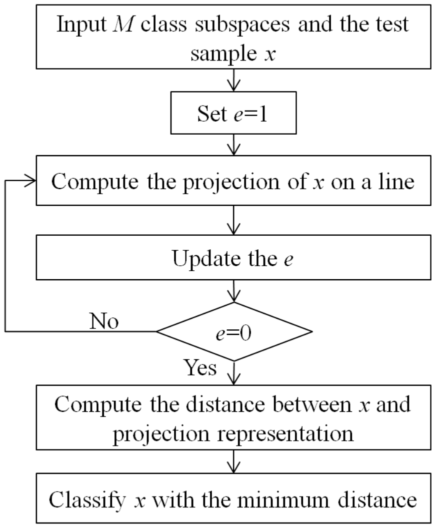

Figure 1: The flowchart of the proposed PRC. $'e = 0'$ means that PRC satisfies the stop condition. The detailed information of e can be found in the stop condition.

Start the iteration For the first iteration, suppose that a class model X_c^0 is described as follows

$$X_c^0 = [x_1^{c,0} \ \ x_2^{c,0} \ \ \cdots \ \ x_{N_c}^{c,0}] \in R^{q \times N_c}, \tag{6}$$

where $x_i^{c,0} = x_i^c, i = 1, 2, \cdots, N_c$. We then select the nearest point $x_{i^*}^{c,0}$ from the class model X_c^0 as follow.

$$i^* = \arg\min(\|x - x_i^{c,0}\|), \ \ i = 1, 2, \cdots, N_c \tag{7}$$

Use the nearest point $x_{i^*}^{c,0}$ and a training sample $x_i^{c,0}$, $i \neq i^*$ to form a line $\overline{x_{i^*}^{c,0} x_i^{c,0}}$. Next, the projection point of the test sample x on the line $\overline{x_{i^*}^{c,0} x_i^{c,0}}$ can be computed by:

$$p_{i^*}^{c,0} = x_{i^*}^{c,0} + t^0 (x_i^{c,0} - x_{i^*}^{c,0}) \tag{8}$$

where $t \in R$ is the position parameter. The vector $\overline{x p_{i^*}^{c,0}}$ is orthogonal to $\overline{x_{i^*}^{c,0} x_i^{c,0}}$, that is, $(x - p_{i^*}^{c,0})(x_i^{c,0} - x_{i^*}^{c,0}) = 0$ where $\bullet$ denotes the dot product. Therefore, the position parameter can be computed as

$$t^0 = \frac{(x - x_{i^*}^{c,0})^T (x_i^{c,0} - x_{i^*}^{c,0})}{(x_i^{c,0} - x_{i^*}^{c,0})^T (x_i^{c,0} - x_{i^*}^{c,0})} \tag{9}$$

For the $(k+1)$th $(k \geq 1)$ iteration, it is easy to know that the projection point $p_{i^*}^{c,k-1}$ in the k^{th} iteration is nearest point. That is, $x_{i^*}^{c,k} = p_{i^*}^{c,k-1}$. The projection point $p_{i^*}^{c,k}$ in the $(k+1)^{th}$ iteration will be computed by x and a line constituted by $p_{i^*}^{c,k-1}$ and another train sample. The procedure of computing the projection point of x on a line is similar to the first iteration.

Rule: All the projection procedure is similar. However, they need to satisfy the following rule. The number of samples from the class subspace is fixed. The new projection points will replace the farther point of the line because they are closer to the test sample. All the samples of class subspace will be sequentially used to constitute the line (as farther point of the line) so that the projection point may contain the information of all the training samples.

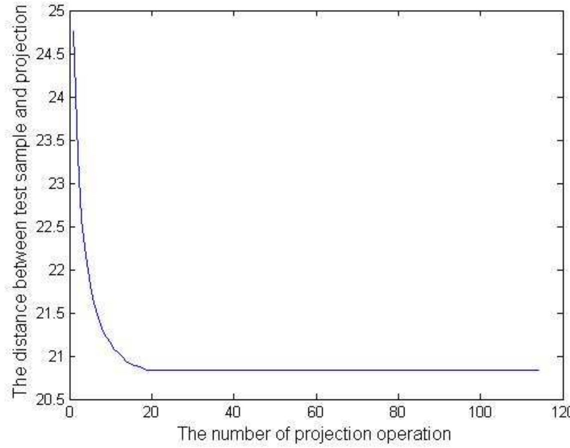

Figure 2: An example of convergence analysis: The distances variations between the test sample and approximation projection point in the iteration procedure.

Convergence Analysis and Stop Condition Because the number of iterations is unlimited, we need to determine the condition for stopping the iteration processes. In order to obtain a good parameter for ending of the process, an example is given as follows. The training set and test sample are produced randomly, the dimension of each sample is 5000, and the number of training samples is 20. Fig. 2 shows that the distance between the test sample and the projection point changes with the number of iterations. As can be seen, the difference between two adjacent distances tends to zeros. Thus, the stop conditions of the iteration process are described as follows.

Condition 1: Suppose that $p_{i^*}^{c,k-1}$ and $p_{i^*}^{c,k}$ are two nearest projection points in the kth and $(k+1)$th iterations. If $\delta < \delta_0$, the iteration process stops, where δ_0 is given before the iteration and the threshold value δ can be computed as

$$\delta = abs\left(\frac{\|x - p_{i^*}^{c,k-1}\| - \|x - p_{i^*}^{c,k}\|}{\|x - p_{i^*}^{c,k-1}\| + \|x - p_{i^*}^{c,k}\|}\right). \tag{10}$$

Besides, in order to avoid the unpredicted situation, another condition is described as follows.

Condition 2: Set the maximum iterative times J. Based on the Figure 2, we suggest that J is set no more than 100. Notice that this condition is rarely used. It can be treated as an insurance.

Set a stop parameter $e = 1$, if one of the two stop conditions is satisfied, $e = 0$, the iteration stops. Then the projection representation p^c can be described as

$$p^c = p_{i^*}^{c,k} = x_{i^*}^{c,k} + t^k (x_i^{c,k} - x_{i^*}^{c,k}) \tag{11}$$

Algorithm 1: Projection Representation

Inputs The entire training samples x_i^c, $c = 1, 2, \cdots, M$, $i = 1, 2, \cdots, N_c$ and a test image vector $x \in R^{q\times 1}$. The stop parameter δ_0 and J.

Output The projection representation p^c.

1. Set $e = 1$; $J = 100$; $\delta_0 = 0.01$
   Repeat
2. Find the nearest point $x_n^{c,k}$ from the class-models X_c^k
3. Compute the projection point $p_{i*}^{c,k}$ of the test sample x on the line $\overline{x_{i*}^{c,k} x_i^{c,k}}$ ($i = 1, 2, \cdots, N_c$ and $i \neq i^*$) as
$$\begin{cases} p_{i*}^{c,k} = x_{i*}^{c,k} + t^k(x_i^{c,k} - x_{i*}^{c,k}) \\ t = \frac{(x-x_{i*}^{c,k})^T(x_i^{c,k}-x_{i*}^{c,k})}{(x_i^{c,k}-x_{i*}^{c,k})^T(x_i^{c,k}-x_{i*}^{c,k})} \end{cases}$$
4. Update the class-models X_c^k using the $p_{i*}^{c,k}$ to replace the farther point $x_i^{c,k}$ of line $\overline{x_{i*}^{c,k} x_i^{c,k}}$ as X_c^{k+1}
5. Update the parameters δ by using the projection point $p_{i*}^{c,k-1}$ in the last iteration and $p_{i*}^{c,k}$ in this iteration as
$$\delta = abs\left(\frac{\|x - p_{i*}^{c,k-1}\| - \|x - p_{i*}^{c,k}\|}{\|x - p_{i*}^{c,k-1}\| + \|x - p_{i*}^{c,k}\|}\right)$$
6. Update the parameter $J = J - 1$.
7. Update the e as
   If ($\delta < \delta_0 \| J < 0$)
   $e = 0$; break;
   end if
   Until the $e = 0$ **and output the** $p^c = p_{i*}^{c,k}$

where

$$t^k = \frac{(x - x_{i*}^{c,k})^T(x_i^{c,k} - x_{i*}^{c,k})}{(x_i^{c,k} - x_{i*}^{c,k})^T(x_i^{c,k} - x_{i*}^{c,k})} \tag{12}$$

Notice: For the example of convergence analysis in Figure 2, we repeat the experiment more than one hundred times. The tendency of the distance variations is similar. Select only some valuable samples that is helpful for classification.

Classification Using the Algorithm 1, the approximation projection p^c is obtained for the c^{th} class subspace. The distance between the test sample and the c^{th} class subspace can be computed as

$$d_c(x) = \|x - p^c\|. \tag{13}$$

PRC selects the class with the minimum distance

$$\min_{c^*} \ d_c(x), c = 1, 2, \cdots, M. \tag{14}$$

Computational Complex

Suppose the dimensional of each sample is q, it is easy to know that the computational cost of each projection operation is $O(q)$. Therefore, the computational complex of PRC is $O(Kq)$, K is the number of projection operations. From the Figure 2, we know that the iteration number is not large, that is to say, the computational cost of PRC is small.

Proposed DPRC

PRC obtains the 'ideal projection' while it doesn't use discriminant analysis for classification. Thus, this section pay attention to utilize the labeled training samples set to constitute a more reliable subspace such that the effective discriminant information can be used for classification. In order to obtain an effective discriminant subspace for PRC, we propose a novel method, called discriminant PRC (DPRC), which obtains the discriminant information by maximizing the ratio of the between-class reconstruction error over the within-class reconstruction error by the PRC.

Optimization of DPRC

The proposed DPRC approach is formulated as the optimization problem to maximize the objective function as,

$$\max_P J(P) = \max_P \frac{J_b}{J_w} \tag{15}$$

where P is the optimal projection matrix that we want to estimate, J_b and J_w denote the between-class and within-class reconstruction representative metrics, respectively. Then, the goal of the DPRC approach becomes to find an optimal mapping matrix, $P = [p_1, ..., p_k, ..., p_d]$ which could project the original sample x_i to a new data sample as $w_i = P^T x_i$ for $i = 1, 2...L$. The proposed projection reduces the dimension and is effective for classification. The above objective function can be also expressed as

$$\begin{aligned} J(P) =& \frac{J_b}{J_w} \\ =& \frac{\frac{1}{L(M-1)} \sum_{i=1}^{L} \sum_{j=1, j\neq l(x_i)}^{M} ||w_i - w_{ij}^b||}{\frac{1}{L}\sum_{i=1}^{L} ||w_i - w_i^w||} \\ =& \frac{\frac{1}{L(M-1)} \sum_{i=1}^{L} \sum_{j=1, j\neq l(x_i)}^{M} ||P^T x_i - P^T x_{ij}^b||}{\frac{1}{L}\sum_{i=1}^{L} ||P^T x_i - P^T x_i^w||} \end{aligned} \tag{16}$$

where $w_{ij}^b = P^T x_{ij}^b$, $w_{ij}^w = P^T x_i^w$, x_{ij}^b and x_i^w are the between-class and within-class projection vectors. That is, they are projection representation of x_i on X_j^b and X_i^w, respectively. They can be calculated by Algorithm 1 with the x_i, X_i^w and X_j^b. X_i^w denotes the $l(x_i)$-th class-model in (1) (don't include the sample x_i), $l(x_i)$ denotes the class label of x_i , X_j^b denotes the j-th ($j \neq l(x_i)$) class-model, With some algebraic derivations in matrices, we have

$$\begin{aligned} J(P) =& \frac{\frac{1}{L} \sum_{i=1}^{L} \sum_{j=1, j\neq l(x_i)}^{M} tr[P^T(x_i - x_{ij}^b)(x_i - x_{ij}^b)^T P]}{\frac{1}{L}\sum_{i=1}^{L} tr[P^T(x_i - x_i^w)(x_i - x_i^w)^T P]} \\ =& tr(\frac{P^T J_b P}{P^T J_w P}) \end{aligned} \tag{17}$$

where

$$J_b = \frac{1}{L}\sum_{i=1}^{L}\sum_{j=1,j\neq l(x_i)}^{M}(x_i - x_{ij}^b)(x_i - x_{ij}^b)^T \quad (18)$$

and

$$J_w = \frac{1}{L}\sum_{i=1}^{L}(x_i - x_i^w)(x_i - x_i^w)^T \quad (19)$$

Afterwards, the objective function can be expressed as

$$\begin{aligned} &\arg\max_P \frac{P^T J_b P}{P^T J_w P} \\ &s.t. \quad P^T P = I \end{aligned} \quad (20)$$

In order to address the typical small sample size problem, the term εI is increased without affecting the subspace. Thus, the objective function can be rewritten as

$$\begin{aligned} &\arg\max_P \frac{P^T J_b P}{P^T (J_w + \varepsilon I) P} \\ &s.t. \quad P^T P = I \end{aligned} \quad (21)$$

where ε is a small number and I is an identity matrix. By using Lagrange multiplier, the projection matrix $P = [p_1, ..., p_k, ..., p_d]$ that maximizes the objective function, which can be gained by solving the eigen decomposition problem of $\frac{J_b}{J_w + \varepsilon I}$ as

$$J_b p_k = \lambda_k (J_w + \varepsilon I) p_k \quad , k = 1, 2, ..., d \quad (22)$$

where $\lambda_1 \geq ... \geq \lambda_k \geq ... \geq \lambda_d$ is d largest eigenvalues and their corresponding eigenvectors, $p_1, ..., p_k, ..., p_d$ of $\frac{J_b}{J_w}$. It is noted that $P = [p_1, ..., p_k, ..., p_d]$ is a $q \times d$ projection matrix, which can project the original q-element data vector to the new d-element data vector as $w_i = P^T x_i$ for $i = 1, 2...L$.

Classification

In the above Section, DPRC obtains the effective discriminant space $W = \{w_i \in R^{d\times 1}, i = 1, 2, \cdots, L\}$. Using the discriminant space W and Algorithm 1, the approximation projection p^c is obtained for the c^{th} class subspace. The distance between the test sample and the c^{th} class subspace can be computed as

$$d_c(w) = \|w - p^c\|. \quad (23)$$

DPRC selects the class with the minimum distance

$$\min_{c^*} d_c(w), c = 1, 2, \cdots, M. \quad (24)$$

where $w = P^T x$.

DPRC vs ULDA

This section compares DPRC with a discriminant-based method: Uncertain LDA (ULDA) (Saeidi, Astudillo, and Kolossa 2016). To better explain it, their similarity and difference are given as follows.

- Similarity: They both maximize the following objective function $\max_P J(P) = \max_P \frac{J_b}{J_w}$, where J_b, J_w are the within-class and between-class scatters. This objective function is the same to that in LDA (Haeb-Umbach and Ney 1992).
- Difference: In ULDA, $J_b = S_b + U_b, J_w = S_w + U_w$, where S_b, S_w are the within-class and between-class scatters in LDA. ULDA proposes the uncertain within-class and between-class scatters U_b, U_w. In DPRC, the J_b, J_w can be treated as new projection-based within-class and between-class scatters, which has significantly difference to S_b, S_w and U_b, U_w.

Experimental Results

This section evaluate the proposed PRC and DPRC on several vision recognition databases.

Face recognition

LFW-a database (Zhu et al. 2012) is used in this experiment. Following (Zhang et al. 2015), we apply 158 subjects that have no less than ten samples for evaluation. The experiment set: 5 samples are randomly selected to form the training set, while other 2 samples are exploited for testing. The SRC (Wright et al. 2009), SVM (Schüldt, Laptev, and Caputo 2004), FDDL (Yang et al. 2014), MCT (Zhang et al. 2015), RCR (Yang et al. 2012), ULDA (Saeidi, Astudillo, and Kolossa 2016), ProCRC (Cai et al. 2016) and CRC (Zhang, Yang, and Feng 2011) algorithms are chosen for comparison. Table 1 illustrates the comparison results of all methods. DPRC obtains better performance than PRC. Compared to the exsiting methods, DPRC has more than 3% improvement..

Classifier	Accuracy	Classifier	Accuracy (%)
SRC	44.10	CRC	44.30
SVM	43.30	ULDA	44.30
FDDL	42.00	ProCRC	44.90
MCT	44.90	**PRC**	**46.84**
RCR	36.70	**DPRC**	**47.90**

Table 1: The recognition rate (RR) of several classifiers on LFW face database

Scene classification

The well-known 15 scene database contains 4,485 images of 15 scene categories (Lazebnik, Schmid, and Ponce 2006). Each image is transformed to spatial pyramid feature provided by (Jiang, Lin, and Davis 2013). The following experimental protocol is used (Liu and Liu 2015): 100 images per class are randomly chosen for training and the rest images are used for testing. The D-KSVD (Zhang and Li 2010), LLC (Wang et al. 2010), LC-KSVD (Jiang, Lin, and Davis 2013), ULDA (Saeidi, Astudillo, and Kolossa 2016), LLNMC (Liu and Liu 2015), LLKNNC (Liu and Liu 2015), LRC (Naseem, Togneri, and Bennamoun 2010), CRC (Zhang, Yang, and Feng 2011), SRC (Wright et al. 2009), ProCRC (Cai et al. 2016), DADL (Guo et al. 2016) methods are chosen for comparison. The average classification rate of 10 runs is used to evaluate all methods. From the results in Table 2, our proposed PRC and DPRC obtain the best performance compared with other methods.

Classifier	Accuracy	Classifier	Accuracy (%)
LRC	95.51	CRC	95.95
LLC	80.57	SSRC	96.45
D-KSVD	89.10	SRC	96.53
LC-KSVD	90.40	ProCRC	96.54
ULDA	97.70	DADL	98.30
LLKNNC	93.54	**PRC**	**98.47**
LLNMC	97.45	**DPRC**	**98.70**

Table 2: The recognition rate (RR) of several classifiers on the 15 scenes database.

Object Classification

The Caltech101 dataset (Fei-Fei, Fergus, and Perona 2007) has 9,144 images with 102 classes. Following the common experimental settings, we train on 5 samples per class and the rest images are used as the testing set. In the experiment, we utilize the 3000-dimension spatial pyramid feature provided by (Jiang, Lin, and Davis 2013) to represent the object image. The DNNC (Zhang et al. 2006), SVM (Schüldt, Laptev, and Caputo 2004), FDDL (Yang et al. 2014), D-KSVD (Zhang and Li 2010), LRC (Naseem, Togneri, and Bennamoun 2010), CRC (Zhang, Yang, and Feng 2011), SRC (Wright et al. 2009), SSRC (Deng, Hu, and Guo 2013), ProCRC (Cai et al. 2016) and ULDA (Saeidi, Astudillo, and Kolossa 2016) methods are chosen for comparison. The experiment results are shown in Table 3. As can be observed, DPRC gains the best performance compared several popular methods.

Classifier	Accuracy	Classifier	Accuracy(%)
DNNC	46.60	SVM	47.88
SRC	48.80	SSRC	47.10
FDDL	49.80	ULDA	48.54
D-KSVD	49.60	ProCRC	47.80
CRC	44.68	**PRC**	**50.66**
LRC	47.54	**DPRC**	**50.80**

Table 3: The recognition rate (RR) of several classifiers on the Caltech 101 database.

Action Recognition

The Ucf50 action dataset (Reddy and Shah 2013) has 6,680 action videos with 50 action categories, which was taken from YouTube. For fair comparison, we follow the ref. (Guo et al. 2016): Divide the database into five folds, use four folds for training and one fold for testing. We use PCA (Luo et al. 2016) to reduce the action bank features (Sadanand and Corso 2012) to 5000 dimensions. The CRC (Zhang, Yang, and Feng 2011), SRC (Wright et al. 2009), DLSI (Ramirez, Sprechmann, and Sapiro 2010), ULDA (Saeidi, Astudillo, and Kolossa 2016), SSRC (Deng, Hu, and Guo 2013) , FDDL (Yang et al. 2014), LC-KSVD (Jiang, Lin, and Davis 2013), DPL (Gu et al. 2014), ProCRC (Cai et al. 2016) and DADL (Guo et al. 2016) methods are chosen for comparison. The experiment results are shown in Table 4. DPRC has the better performance than PRC and gains the best performance compared with several popular methods.

Classifier	Accuracy	Classifier	Accuracy(%)
CRC	75.60	DPL	77.40
SSRC	76.40	ULDA	77.60
SRC	75.00	ProCRC	77.40
DLSI	75.40	DADL	78.00
FDDL	76.50	**PRC**	**78.50**
LC-KSVD	70.10	**DPRC**	**79.10**

Table 4: The recognition rate (RR) of several classifiers on the Ucf50 action database.

Compare with Deep Learning based Methods

The Caltech-256 dataset (Griffin, Holub, and Perona 2007) has 30,608 object images of 256 object class, each class has at least 80 object images. To access the performance of PRC and DPRC for object recognition with the deep-learning-based feature, we follow Ref. (Simon and Rodner 2015), randomly select 60 images for training, the rest images are used for testing. Five deep learning based methods are used for comparison. They include NAC (Simon and Rodner 2015), CNN-S (Chatfield et al. 2014), ZF (Zeiler and Fergus 2014), CNN-M (Chatfield et al. 2014) and VGG19 (Simonyan and Zisserman 2014). The experiment results are shown in Table 5. As we can see, the proposed methods with deep feature obtain the better performance than the deep learning based methods. The proposed DPRC has the better performance compared to proposed PRC.

Classification Methods	Accuracy (%)
CNN-S	77.6
ZF	74.2
CNN-M	75.5
VGG19+ SVM	83.9
NAC	84.1
VGG19+PRC	**84.9**
VGG19+DPRC	**85.3**

Table 5: Accuracy of several methods on the Caltech 256 object database.

Conclusion

In this paper, projection representation-based classification (PRC) has been proposed for image recognition. The PRC uses the iteratively projection procedures to obtain a point to closely approximate the 'ideal representation'. The objectives of PRC, SRC and LRC are similar but PRC gains the better representation. Based on PRC, the discriminant PRC (DPRC) is further proposed. DPRC increase the discriminant information for PRC such that it obtains the better performance. The experimental results on several well-known databases have confirmed the good performance of the proposed PRC and DPRC for face, objection, scene and action recognitions. Moreover, PRC and DPRC with deep-learning-based feature can obtain the better performance than deep learning based methods

Acknowledgments

Thanks for the valuable suggestions of Editor and reviewers. This work was supported in part by the Macau Science and Technology Development Fund under Grant FDCT/016/2015/A1 and by the Research Committee at University of Macau under Grants MYRG2014-00003-FST and MYRG2016-00123-FST.

References

[Basri and Jacobs 2003] Basri, R., and Jacobs, D. W. 2003. Lambertian reflectance and linear subspaces. *TPAMI* 25(2):218–233.

[Cai et al. 2016] Cai, S.; Zhang, L.; Zuo, W.; and Feng, X. 2016. A probabilistic collaborative representation based approach for pattern classification. In *CVPR*, 2950–2959.

[Chatfield et al. 2014] Chatfield, K.; Simonyan, K.; Vedaldi, A.; and Zisserman, A. 2014. Return of the devil in the details: Delving deep into convolutional nets. In *BMVC*, 1–12.

[Deng, Hu, and Guo 2013] Deng, W.; Hu, J.; and Guo, J. 2013. In defense of sparsity based face recognition. In *CVPR*, 399–406. IEEE.

[Fei-Fei, Fergus, and Perona 2007] Fei-Fei, L.; Fergus, R.; and Perona, P. 2007. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding* 106(1):59–70.

[Feng and Zhou 2016a] Feng, Q., and Zhou, Y. 2016a. Iterative linear regression classification for image recognition. In *ICASSP*, 1566–1570.

[Feng and Zhou 2016b] Feng, Q., and Zhou, Y. 2016b. Kernel combined sparse representation for disease recognition. *TMM* 18(10):1956–1968.

[Feng and Zhou 2017] Feng, Q., and Zhou, Y. 2017. Kernel regularized data uncertainty for action recognition. *TCSVT* 27(3):577–588.

[Feng, Zhou, and Lan 2016] Feng, Q.; Zhou, Y.; and Lan, R. 2016. Pairwise linear regression classification for image set retrieval. In *CVPR*, 4865–4872.

[Griffin, Holub, and Perona 2007] Griffin, G.; Holub, A.; and Perona, P. 2007. Caltech-256 object category dataset.

[Gu et al. 2014] Gu, S.; Zhang, L.; Zuo, W.; and Feng, X. 2014. Projective dictionary pair learning for pattern classification. In *NIPS*, 793–801.

[Guo et al. 2016] Guo, J.; Guo, Y.; Kong, X.; Zhang, M.; and He, R. 2016. Discriminative analysis dictionary learning. In *AAAI*, 1617–1623.

[Haeb-Umbach and Ney 1992] Haeb-Umbach, R., and Ney, H. 1992. Linear discriminant analysis for improved large vocabulary continuous speech recognition. In *ICASSP*, volume 1, 13–16.

[Jiang, Lin, and Davis 2013] Jiang, Z.; Lin, Z.; and Davis, L. S. 2013. Label consistent k-svd: Learning a discriminative dictionary for recognition. *TPAMI* 35(11):2651–2664.

[Lazebnik, Schmid, and Ponce 2006] Lazebnik, S.; Schmid, C.; and Ponce, J. 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*, volume 2, 2169–2178. IEEE.

[Liu and Liu 2015] Liu, Q., and Liu, C. 2015. A novel locally linear knn model for visual recognition. In *CVPR*, 1329–1337.

[Luo et al. 2016] Luo, M.; Nie, F.; Chang, X.; Yang, Y.; Hauptmann, A.; and Zheng, Q. 2016. Avoiding optimal mean robust pca/2dpca with non-greedy l1-norm maximization. In *IJCAI*, 1802–1808.

[Naseem, Togneri, and Bennamoun 2010] Naseem, I.; Togneri, R.; and Bennamoun, M. 2010. Linear regression for face recognition. *TPAMI* 32(11):2106–2112.

[Ramirez, Sprechmann, and Sapiro 2010] Ramirez, I.; Sprechmann, P.; and Sapiro, G. 2010. Classification and clustering via dictionary learning with structured incoherence and shared features. In *CVPR*, 3501–3508. IEEE.

[Reddy and Shah 2013] Reddy, K. K., and Shah, M. 2013. Recognizing 50 human action categories of web videos. *Machine Vision and Applications* 24(5):971–981.

[Sadanand and Corso 2012] Sadanand, S., and Corso, J. J. 2012. Action bank: A high-level representation of activity in video. In *CVPR*, 1234–1241.

[Saeidi, Astudillo, and Kolossa 2016] Saeidi, R.; Astudillo, R. F.; and Kolossa, D. 2016. Uncertain lda: Including observation uncertainties in discriminative transforms. *TPAMI* 38(7):1479–1488.

[Schüldt, Laptev, and Caputo 2004] Schüldt, C.; Laptev, I.; and Caputo, B. 2004. Recognizing human actions: a local svm approach. In *ICPR*, volume 3, 32–36.

[Simon and Rodner 2015] Simon, M., and Rodner, E. 2015. Neural activation constellations: Unsupervised part model discovery with convolutional networks. In *ICCV*, 1143–1151.

[Simonyan and Zisserman 2014] Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*.

[Wang et al. 2010] Wang, J.; Yang, J.; Yu, K.; Lv, F.; Huang, T.; and Gong, Y. 2010. Locality-constrained linear coding for image classification. In *CVPR*, 3360–3367.

[Wright et al. 2009] Wright, J.; Yang, A. Y.; Ganesh, A.; Sastry, S. S.; and Ma, Y. 2009. Robust face recognition via sparse representation. *TPAMI* 31(2):210–227.

[Yang et al. 2012] Yang, M.; Zhang, L.; Zhang, D.; and Wang, S. 2012. Relaxed collaborative representation for pattern classification. In *CVPR*, 2224–2231.

[Yang et al. 2014] Yang, M.; Zhang, L.; Feng, X.; and Zhang, D. 2014. Sparse representation based fisher discrimination dictionary learning for image classification. *IJCV* 109(3):209–232.

[Zeiler and Fergus 2014] Zeiler, M. D., and Fergus, R. 2014. Visualizing and understanding convolutional networks. In *ECCV*, 818–833.

[Zhang and Li 2010] Zhang, Q., and Li, B. 2010. Discriminative k-svd for dictionary learning in face recognition. In *CVPR*, 2691–2698.

[Zhang et al. 2006] Zhang, H.; Berg, A. C.; Maire, M.; and Malik, J. 2006. Svm-knn: Discriminative nearest neighbor classification for visual category recognition. In *CVPR*, volume 2, 2126–2136.

[Zhang et al. 2015] Zhang, B.; Perina, A.; Murino, V.; and Del Bue, A. 2015. Sparse representation classification with manifold constraints transfer. In *CVPR*, 4557–4565.

[Zhang, Yang, and Feng 2011] Zhang, L.; Yang, M.; and Feng, X. 2011. Sparse representation or collaborative representation: Which helps face recognition? In *ICCV*, 471–478.

[Zhu et al. 2012] Zhu, P.; Zhang, L.; Hu, Q.; and Shiu, S. C. 2012. Multi-scale patch based collaborative representation for face recognition with margin distribution optimization. In *ECCV*. 822–835.