FACE RECOGNITION USING EXTENDED GENERALIZED RAYLEIGH QUOTIENT

Jingyi Yang, Chun Qi, Yuhua Li, Jie Li

School of Electronic and Information Engineering
Xi'an Jiaotong University, Xi'an, China, 710049
yangjingyi@stu.xjtu.edu.cn, qichun@mail.xjtu.edu.cn, lyhresearch@njau.edu.cn, jielixjtu@xjtu.edu.cn

ABSTRACT

Generalized Rayleigh quotient is a powerful mathematical tool. This framework can combine two conflicting objectives, the maximization and minimization, in one unified function. Many problems in machine learning can be considered as the optimization of generalized Rayleigh quotient. In this paper, we propose an extension of generalized Rayleigh quotient framework and develop a new method for face recognition based on this framework. This method minimizes the residual of within-class collaborative representation and maximizes the residual of between-class collaborative representation. Then intra-class and inter-class adjacency graphs are constructed as constraints imposed on the two residuals respectively to preserve the consistency of distance property. Solution is iteratively obtained from generalized eigenvalue problem. The proposed method is evaluated on benchmark face databases and outperforms other state-of-the-art methods.

Index Terms— The extension of generalized Rayleigh quotient, Face recognition, Collaborative representation, Graph constraint

1. INTRODUCTION

Rayleigh quotient is a powerful tool in mathematical problems proposed by Rayleigh L [1]. Researches about its mathematical properties have been investigated by Golub and Loan [2]. Generalized Rayleigh quotient was presented by Mead D J. [3]. It is uncommon to directly explore the use of generalized Rayleigh quotient in machine learning field. The framework of generalized Rayleigh quotient can combine two conflicting objectives, the maximization and the minimization, in one unified function. Many problems in machine learning are involved in maximization and the minimization simultaneously. P. N. Belhumeur et al. [4] proposed Fisherface based on linear discriminant analysis (LDA). Such supervised method is formulated by maximizing the measure of between-class scatter while minimizing the measure of within-class scatter. Yang et al. [5] proposed a kernel Fisher discriminant analysis which achieves the same goal as Fisherface in a Hilbert space $\mathcal{H}$. Yang et al. [6] proposed an unsupervised discriminant projection (UDP) to minimize the local scatter and maximize the non-local scatter simultaneously. Local and non-local scatter are defined referring to δ-neighborhoods. Yan et al. [7] proposed the Marginal Fisher Analysis (MFA). MFA minimizes the intra-class compactness by intrinsic graph and maximizes the inter-class separability by a penalty graph. Liu et al. [8] presented constrained maximum variance mapping (CMVM) method, which constructs graphs with class information. CMVM minimizes the local scatter and maximizes the dissimilarity between manifolds. Yang et al. proposed the multi-manifold discriminant analysis (MMDA) [8] which maximizes the between-class Laplacian matrix and minimizes the within-class Laplacian matrix. Yang et al. [9] proposed SRC steered discriminative projection (SRC-DP). SRC-DP minimizes average within-class distance and maximizes the between-class distance defined by l_1-norm. Wright J et al. [10] proposed the subspace learning with l_1-graph. They modified linear discriminant analysis (LDA) by imposing the constraint of l_1-graph, thus valuable information for classification can be preserved after projection. Huang et al. [11] proposed collaborative discriminant locality preserving projections (CDLPP) to minimize the Laplacian matrix of within-class and maximize the scatter matrix of between-class. They imposed a l_2-norm constraint on the projections to discover the collaboration relationship between dimensions.

The methods mentioned above are the problems that require maximization and minimization simultaneously, so these problems can be represented by generalized Rayleigh quotient. In addition, some of these methods such as Fisherface [4], kernel LDA [5], UDP [6], MFA [7], CMVM [8], MMDA [8] and SRC-DP [9] have no constraint while other methods such as LDA with l_1-norm constraint [10], CDLPP [11] have constraints so as to obtain desiring performance. We propose an extension of generalized Rayleigh quotient. It not only minimizes one object and maximizes another object but also imposes constraints on maximization or minimization part. The methods mentioned above can be represented by proposed generalized Rayleigh quotient framework.

This research is supported by the National Natural Science Foundation of China (Grant No. 61572395 and 61675161).

978-1-5090-6067-2/17/$31.00 ©2017 IEEE

Based on proposed extended framework, we present a new discriminative method for face recognition. Using the residuals of collaborative representation and graph constraint, we then compute a transformation matrix projecting data into a low-dimensional space. Recognition is implemented by comparing the class specific representation residuals in this low-dimensional space.

The rest of this paper is organized as follows. Section 2 reviews generalized Rayleigh quotient and proposes a framework based on the extension of generalized Rayleigh quotient. Our method is proposed in Section 3. Experiments and results are presented in Section 4. Conclusions are given in Section 5.

2. REVIEW OF GENERALIZED RAYLEIGH QUOTIENT AND RELATED FRAMEWORK

Rayleigh quotient is a powerful tool in mathematical problems proposed by Rayleigh L [1]. Generalized Rayleigh quotient was presented by Mead D J. [3]. It is formulated as:

$$J = \frac{\mathbf{p^T A p}}{\mathbf{p^T B p}} \tag{1}$$

where $\mathbf{A}$ and $\mathbf{B}$ are positive definite matrices, $\mathbf{p}$ is a vector and $\mathbf{p} \neq \mathbf{0}$. The common goal is to minimize or maximize J. With the condition $\mathbf{p^T B p} = 1$, $\mathbf{p}$ can be obtained by eigenvector associated with the minimum or maximum eigenvalue of Eq.(2):

$$\mathbf{Ap} = \lambda \mathbf{Bp} \tag{2}$$

If the vector $\mathbf{p}$ is replaced by the matrix $\mathbf{P} \in \mathbf{R}^{m \times d}$, trace should be used to obtain Rayleigh quotient. Thus we can formulate the generalized Rayleigh quotient as Eq.(3).

$$J = \frac{tr(\mathbf{P^T A P})}{tr(\mathbf{P^T B P})} \tag{3}$$

Matrix $\mathbf{P}$ can be obtained from eigenvectors associated with d maximum or minimum eigenvalues by Eq.(2).

Considering specific constraints for specific problems, we propose the extension of generalized Rayleigh quotient shown as Eq.(4) by imposing constraints on denominator and numerator of (3):

$$J = \frac{tr(\mathbf{P}^T \mathbf{AP} + \lambda_1 \mathbf{C}_1 + \cdots + \lambda_n \mathbf{C}_n)}{tr(\mathbf{P}^T \mathbf{BP} + \mu_1 \mathbf{D}_1 + \cdots + \mu_m \mathbf{D}_m)} \tag{4}$$

Assuming that $\mathbf{C}_i (i = 1, \cdots, n)$ and $\mathbf{D}_j (j = 1, \cdots, m)$ are constraints, $\lambda_i (i = 1, \cdots, n)$ and $\mu_j (j = 1, \cdots, m)$ are parameters to balance the constraints respectively. By using this framework, we can not only maximize one object and minimize another one but also impose constraints on maximization or minimization part. Based on this framework, we propose a method for face recognition. It serves as an example of how to use the extended Rayleigh quotient.

3. OUR PROPOSED METHOD BASED ON EXTENDED GENERALIZED RAYLEIGH QUOTIENT

In this section, we give details of the proposed method. For discriminative classification, we impose two kinds of graph constraints on two collaborative representation residuals respectively. Then maximization and minimization are realized simultaneously from proposed extended framework.

3.1. Collaborative representation residual

A set of face image samples $\{\mathbf{x}_i\}$ belonging to c classes can be represented as a $m \times n$ matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \in \mathbf{R}^{m \times n}$. By projecting $\mathbf{x}_i$ onto $\mathbf{y}_i$ via $\mathbf{y}_i = \mathbf{P}^T \mathbf{x}_i$, where $\mathbf{P} \in \mathbf{R}^{m \times d}$ ($d \ll m$) is a transformation matrix, we obtain the projected matrix $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_n] \in \mathbf{R}^{d \times n}$ in the low-dimensional space. Coding every image $\mathbf{y}_i$ ($i = 1, \cdots, n$) over the dictionary $\mathbf{Y}_{dict} = [\mathbf{y}_1, \cdots, \mathbf{y}_{i-1}, \cdots, \mathbf{y}_{i+1}, \mathbf{y}_n] \in \mathbf{R}^{d \times (n-1)}$, we get a set of collaborative representation coefficients $\{\mathbf{w}_i\}_{i=1}^n$ by

$$\mathbf{w_i} = \arg\min_{\mathbf{w}_i} \|\mathbf{y}_i - \mathbf{Y}\mathbf{w}_i\|_2^2 \quad s.t. w_{ii} = 0 \tag{5}$$

Assume that the sample $\mathbf{y}_i$ belongs to g^{th} class, $\delta_g(w_i)$ means that only the coefficients associated with samples from g^{th} class are remained while other coefficients are set zero. Representation residuals can be divided into two categories. The first type is the residual of representation by face samples in the same class, which defined as $r_i = \|\mathbf{y}_i - \mathbf{Y}\delta_g(\mathbf{w}_i)\|_2^2$. r_i arises from expression and illumination variation. The second type is residual of representation by face samples of different classes. It is defined as $\bar{r_i} = \|\mathbf{y}_i - \sum_{s=1, s\neq g}^{c} \mathbf{Y}\delta_s(\mathbf{w}_i)\|_2^2$. $\bar{r_i}$ results from subject identity difference. Obviously, r_i should be as small as possible and $\bar{r_i}$ should be as large as possible to increase the discrimination of projected space.

Considering all the samples, we define a within-class representation residual and a between-class representation residual as Eq.(6) and (8). The detailed derivations of Eq.(6) and (8) can be found in supplemental material.

$$r_s = \frac{1}{n} \sum_{i=1}^{n} \|\mathbf{y}_i - \mathbf{Y}\delta_g(\mathbf{w}_i)\|_2^2 = tr(\mathbf{P}^T \mathbf{R}_s \mathbf{P}) \tag{6}$$

where

$$\mathbf{R}_s = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{X}\delta_g(\mathbf{w}_i))(\mathbf{x}_i - \mathbf{X}\delta_g(\mathbf{w}_i))^T \tag{7}$$

$$r_d = \frac{1}{n} \sum_{i=1}^{n} \|\mathbf{y}_i - \sum_{s=1, s\neq g}^{c} \mathbf{Y}\delta_s(\mathbf{w}_i)\|_2^2 = tr(\mathbf{P}^T \mathbf{R}_d \mathbf{P}) \tag{8}$$

where

$$\mathbf{R}_d = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i - \sum_{s=1, s\neq g}^{c} \mathbf{X}\delta_s(\mathbf{w}_i))(\mathbf{x}_i - \sum_{s=1, s\neq g}^{c} \mathbf{X}\delta_s(\mathbf{w}_i))^T \tag{9}$$

In order for an accurate classification, r_s should be minimized while r_d is maximized. Therefore, discriminative information hidden in the training samples are well exploit.

3.2. Graph constraint

In the low-dimensional space, we expect samples of same class stay as close as possible and samples of different classes stay as far as possible. To satisfy such criterion, we define two types of graphs, intra-class graph and inter-class graph. For intra-class graph, an edge is constructed between nodes $\mathbf{x}_i$ and $\mathbf{x}_j$ which are from the same class. The adjacency matrix $\mathbf{H}$ for intra-class graph is defined as:

$$H_{ij} = \begin{cases} 1 & \text{if class}(x_i)\text{=class}(x_j) \\ 0 & \text{else} \end{cases} \tag{10}$$

We minimize

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\mathbf{y}_i - \mathbf{y}_j\|_2^2 H_{ij} = tr(\mathbf{P}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{P}) \tag{11}$$

where $\mathbf{L} = \mathbf{D} - \mathbf{H}$ is the Laplacian matrix [12]; $\mathbf{D}$ is a diagonal matrix, and its entries are column sum of $\mathbf{H}$, that is $D_{ii} = \sum_j H_{ij}$. The detailed derivation of Eq.(11) can be found in supplemental material. By minimizing Eq.(11), samples in the same class stay close to each other in the low-dimensional space. Then we consider the distribution of samples of different classes. Supposing that $\mathbf{x}_i^g$ and $\mathbf{y}_i^g$ are samples of g^{th} class and the number of samples belonging to g^{th} class is n_g, we compute the low-dimensional mean face and mean face in original space:

$$\mathbf{m}_g = \frac{1}{n_g}\sum_{i=1}^{n_g}\mathbf{y}_i^g \tag{12}$$

$$\mathbf{f}_g = \frac{1}{n_g}\sum_{i=1}^{n_g}\mathbf{x}_i^g \tag{13}$$

where $g = 1, 2, \cdots, c$. Adjacency matrix $\mathbf{B}$ for inter-class graph is determined by $B_{gs} = exp(-\|\mathbf{m}_g - \mathbf{m}_s\|_2^2/t)$, where t is an empirically determined parameter. To remain samples of different classes stay as far as possible, we maximize:

$$\begin{aligned}
&\frac{1}{2}\sum_{g=1}^{c}\sum_{s=1}^{c}\|\mathbf{m}_g - \mathbf{m}_s\|_2^2 B_{gs} \\
=&tr(\sum_{g=1}^{c}\mathbf{P}^T\mathbf{f}_g E_{gg}\mathbf{f}_g^T\mathbf{P} - \sum_{g=1}^{c}\sum_{s=1}^{c}\mathbf{P}^T\mathbf{f}_g B_{gs}\mathbf{f}_s^T\mathbf{P}) \\
=&tr(\mathbf{P}^T\mathbf{F}\mathbf{Z}\mathbf{F}^T\mathbf{P})
\end{aligned} \tag{14}$$

where $\mathbf{Z} = \mathbf{E} - \mathbf{B}$; $\mathbf{E}$ is a diagonal matrix and $E_{gg} = \sum_s B_{gs}$; $\mathbf{F} = [\mathbf{f}_1, \cdots, \mathbf{f}_n]$. Detailed derivation is shown in supplemental material. In this step, we calculate mean faces instead of directly using individual samples. Thus the influence of random noise and expression, illumination variation on the distribution of samples will be abated. By the two graphs we ensure that the similar samples are mapped near to each other and the dissimilar samples are mapped far away from each other in the low-dimensional space. Compared with manifold learning methods such as locality preserving projections(LPP), constrained maximum variance mapping(CMVM), label information is used and the number of nearest neighbors does not need to be chosen.

3.3. Objective function

With the intra-class and inter-class graph constraints imposing on within-class and between-class representation residuals respectively, our objective function is represented in extended generalized Rayleigh quotient framework (4):

$$\begin{aligned}
\max J =&\frac{tr(\mathbf{P}^T\mathbf{R}_d\mathbf{P} + \lambda_1\mathbf{P}^T\mathbf{F}\mathbf{Z}\mathbf{F}^T\mathbf{P})}{tr(\mathbf{P}^T\mathbf{R}_s\mathbf{P} + \mu_1\mathbf{P}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{P})} \\
=&\frac{tr(\mathbf{P}^T\mathbf{U}\mathbf{P})}{tr(\mathbf{P}^T\mathbf{T}\mathbf{P})}
\end{aligned} \tag{15}$$

where $\mathbf{U} = \mathbf{R}_d + \lambda_1\mathbf{F}\mathbf{Z}\mathbf{F}^T$, $\mathbf{T} = \mathbf{R}_s + \mu_1\mathbf{X}\mathbf{L}\mathbf{X}^T$, λ_1 and μ_1 are positive parameters.

Then we compute the eigenvectors and eigenvalues for the generalized eigenvalue problem: $\mathbf{U}\varphi = \gamma\mathbf{T}\varphi$. Select the column vectors $\varphi_1, \cdots, \varphi_d$ associated with the largest d eigenvalues $\gamma_1, \cdots, \gamma_d$ to be the solution of Eq.(15). Obtaining projection matrix, we map samples with the new projection and repeat the steps above until the next value of J is smaller or equal to the older one. Our experiments illustrate that the performance of recognition is insensitive to the initiation of projection. So we initialize $\mathbf{P}$ with a random matrix and iteratively find the optimal projection matrix $\mathbf{P}^*$.

For a query sample $\mathbf{x}$, it is mapped into the low-dimensional space by the optimal projection matrix $\mathbf{P}^*$ with $\mathbf{y} = \mathbf{P}^{*T}\mathbf{x}$. We code $\mathbf{y}$ over training set $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_n] \in \mathbf{R}^{d\times(n-1)}$ by $\alpha = \arg\min_\alpha \|\mathbf{y} - \mathbf{Y}\alpha\|_2^2, \alpha_i = 0$. The identity of the query image could be estimated by the smallest coding residual as Eq.(16). The proposed method is summarized in Algorithm 1.

$$\text{identity}(\mathbf{x}) = \arg\min_g \|\mathbf{y} - \mathbf{Y}\delta_g(\alpha)\|_2^2/\|\delta_g(\alpha)\|_2^2 \tag{16}$$

3.4. Geometric explanation of proposed method

In this section, we will illustrate how our algorithm works by analyzing the meaning of objective function in geometric perspective in Fig.1. Given a mapped sample $\mathbf{y}_i$ which belongs to g^{th} class, we code it over dictionary $\mathbf{Y}_{dict} = [\mathbf{y}_1, \cdots, \mathbf{y}_{i-1}, \mathbf{y}_{i+1}, \mathbf{y}_n]$. It is obvious that $\mathbf{Y}\mathbf{w}_i = \mathbf{Y}\delta_g(\mathbf{w}_i) + \sum_{s\neq g}\mathbf{Y}\delta_s(\mathbf{w}_i)$. Denote $\mathbf{Y}\delta_g(\mathbf{w}_i)$ by γ_g and $\sum_{s\neq g}\mathbf{Y}\delta_s(\mathbf{w}_i)$ by $\bar{\gamma}_g$, shown in Fig.1. Thus we have $\|\mathbf{y}_i -$

$\gamma_g\|_2^2 = \|\mathbf{y}_i - \mathbf{Y}\mathbf{w}_i\|_2^2 + \|\mathbf{Y}\mathbf{w}_i - \gamma_g\|_2^2$ and $\|\mathbf{y}_i - \bar{\gamma_\mathbf{g}}\|_2^2 = \|\mathbf{y}_i - \mathbf{Y}\mathbf{w}_i\|_2^2 + \|\mathbf{Y}\mathbf{w}_i - \bar{\gamma_\mathbf{g}}\|_2^2$. Ignoring the role of part $\|\mathbf{y}_i - \mathbf{Y}\mathbf{w}_i\|_2^2$ for classification, the following equation is induced:

$$\frac{\|\mathbf{Y}\mathbf{w}_i - \gamma_g\|_2^2}{\|\mathbf{Y}\mathbf{w}_i - \bar{\gamma_\mathbf{g}}\|_2^2} = \frac{sin^2(\mathbf{Y}\mathbf{w}_i, \gamma_g)}{sin^2(\mathbf{Y}\mathbf{w}_i, \bar{\gamma}_g)} \tag{17}$$

So Eq.(17) indicates that maximizing $\|\mathbf{y}_i - \gamma_g\|_2^2$ and minimizing $\|\mathbf{y}_i - \bar{\gamma_\mathbf{g}}\|_2^2$ lead to a discriminant angle relationship. If this criteria is applied to every sample, the classification will be more accurate because of increasing discrimination.

The within-class representation residual defined in section 3.1 can be written as:

$$\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{y_i} - \mathbf{Y}\delta_g(\mathbf{w_i})\|_2^2 = \frac{1}{n}\sum_{i=1}^{n}\|\mathbf{y_i} - \gamma_\mathbf{g}\|_2^2 \tag{18}$$

Equivalently, the between-class representation residual is:

$$\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{y_i} - \sum_{s=1, s\neq g}^{c}\mathbf{Y}\delta_s(\mathbf{w}_i)\|_2^2 = \frac{1}{n}\sum_{i=1}^{n}\|\mathbf{y_i} - \bar{\gamma_\mathbf{g}}\|_2^2 \tag{19}$$

Thus with within-class and between-class residual, we find a discriminative angle relationship. In addition, to enhance the discrimination, the distance property of within-class samples and between-class samples need to be optimized, thus we impose two graph constraints on residuals respectively. Obviously, by minimizing within-class representation residual and maximizing between-class representation residual, the goal of minimizing the difference arising from expression and illumination variation as well as maximizing the distinctiveness between different subjects could be achieved simultaneously. Using graph constraint, requirements for distances between samples from same or different classes are directly satisfied, which enhance the discrimination of recognition.

4. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of proposed method by using publicly available database for face recognition. We will first demonstrate the recognition performance on AR database [13] and Extended Yale B database [14]. Then we demonstrate its robustness to occlusion by using AR disguised database and Extended Yale B database with artificial occlusion. For proposed method, we set parameters $\lambda_1 = 10$ and $\mu_1 = 1$, threshold $\varepsilon = 10^{-6}$, parameter $t = 0.01$ for inter-class graph.

Samples are preprocessed by PCA to dimension 300 and they are used as the input facial features for all the methods.

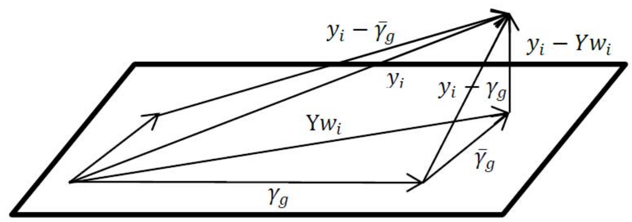

Fig. 1. The geometric explanation of proposed method

Algorithm 1 Our Proposed Method

Training process

Input: $[\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \in \mathbf{R}^{m\times n}$, query image $\mathbf{x} \in \mathbf{R}^{m\times 1}$, class number c, parameter λ_1, μ_1, t, threshold ε, objective dimension number d

Initialize: $\mathbf{J}_0$, $\mathbf{P}$(initialize with a random matrix)

Output: Identity of query image $\mathbf{x}$

WHILE $(J - J_0)/J_0 > \varepsilon$ **do**

1.Map m-dimensional Training set $\{\mathbf{x}_i\}$ to d-dimensional space($d \ll m$) by $\mathbf{y}_i = \mathbf{P}^T\mathbf{x}_i, \mathbf{P} \in \mathbf{R}^{m\times d}, i = 1, \cdots, n.$

2. Calculate collaborative representation coefficients by
$$\mathbf{w}_i = \arg\min_{\mathbf{w}_i} \|\mathbf{y}_i - \mathbf{Y}\mathbf{w}_i\|_2^2 \quad s.t. w_{ii} = 0$$

3. Calculate collaborative representation residual matrices $\mathbf{R}_s$ and $\mathbf{R}_d$ by Equ.(7) and (9).

4. Calculate matrix $\mathbf{L}$ and $\mathbf{Z}$ by $\mathbf{L} = \mathbf{D} - \mathbf{H} \quad \mathbf{Z} = \mathbf{E} - \mathbf{B}$.

5. Obtain the matrix $\mathbf{U}$ and $\mathbf{T}$ by
$$\mathbf{U} = \mathbf{R}_d + \lambda_1\mathbf{F}\mathbf{Z}\mathbf{F}^T$$
$$\mathbf{T} = \mathbf{R}_s + \mu_1\mathbf{X}\mathbf{L}\mathbf{X}^T$$
Solve the generalized eigenvalue problem $\mathbf{U}\varphi = \gamma\mathbf{T}\varphi$, get projection matrix $\mathbf{P}$ then update $\mathbf{P}$.

6. Calculate the value of objective function J.

END WHILE

Recognition process

1. Map $\mathbf{x}$ into the d-dimensional space with optimal projection matrix $\mathbf{P}^*$ which is obtained during training process:
$$\mathbf{y} = \mathbf{P}^{*T}\mathbf{x}$$

2. Code $\mathbf{y}$ by:
$$\alpha = \arg\min_{\alpha} \|\mathbf{y} - \mathbf{Y}\alpha\|_2^2$$

3. Compute the regularized residuals by:
$$r_g = \frac{\|\mathbf{y} - \mathbf{Y}\delta_g(\alpha)\|_2^2}{\|\delta_g(\alpha)\|_2^2}, \quad g = 1, 2, \cdots, c$$

4. Output the identity of $\mathbf{x}$ as:
$$identity(\mathbf{x}) = \arg\min_{g}\{r_g\}$$

4.1. Experiments on the AR database

The AR database consists of over 4,000 frontal images from 126 individuals including more facial variations, illumination changes, expressions and occlusions. In the experiment, we

Table 1. Recognition rate of each method for Extended Yale B database

Methods	Our method	CRC	SRC	SRC-DP	SPP	LDA	CMVM	UDP
Recognition rate	**0.987**	0.970	0.971	0.9801	0.952	0.864	0.954	0.947
Dimension	40	180	180	180	180	37	180	180

choose a subset that contains 70 male subjects and 56 female subjects including only illumination and expression changes.

For each subject, we choose seven images for training and another seven for testing. The images are cropped with dimension 60×43. Comparison with CRC [15], SRC [16], SRC-DP [9], CMVM [8], UDP [6], CDLPP [11] is given in Fig.2. Denote that CRC classifier is adopted to predict the class labels of test data for CMVM, UDP, CDLPP. PCA is used to extract features for CRC and SRC. From this figure, we can see that the performance of proposed method exceeds the performance of other methods at each individual feature dimension. To demonstrate the effectiveness of graph constraint, we remove the graph constraint in our method and evaluate the recognition performance. From the line shown in black in Fig.2, the recognition rate without constraint is lower than result obtained with constraint, we can observe that especially in low-dimensional case (lower than 100), graph constraint helps a lot to increase recognition rate. Therefore, graph constraint does play an important role in increasing discrimination of result.

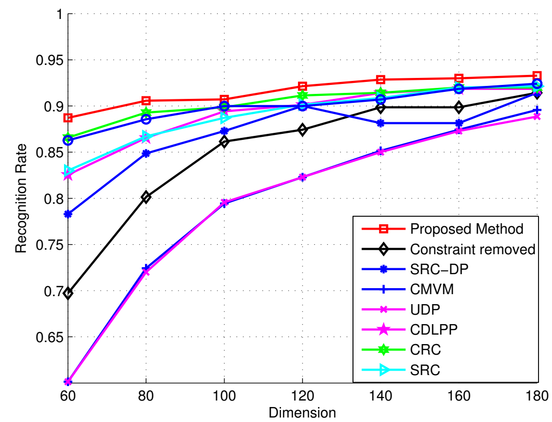

Fig. 2. Recognition rate curve of each method versus variation of dimension (AR database)

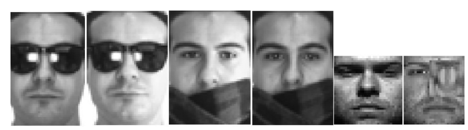

Fig. 3. Examples of images with real and artificial occlusion

4.2. Experiments on the Extended Yale Face database B

In this experiment, the images are cropped and resized to 32x32 pixels. This dataset has 38 individuals and around 64 near frontal images under different illuminations per individual. We randomly split the database into two halves. 32 images are used for training and the remaining ones are used for testing. The results of our method, CRC [15], SRC [16], SRC-DP [9], SPP [17], LDA [4], CMVM [8], UDP [6] are shown in Table.1. Denote that CRC classifier is adopted to predict the class labels of test data for LDA, SPP, CMVM and UDP. PCA is used to extract features for CRC and SRC.

It can be seen from Table.1 that the recognition rate of our method is higher than other listed methods. And our method obtains better result in lower dimension, which confirms that the proposed method has persuasive performance.

4.3. Experiments with occlusion

To evaluate the robustness of our proposed method, we implement the experiments on AR database with real occlusion and Extended Yale B database with artificial occlusion.

Disguise in the AR database

In this experiment, we use one image with sunglasses, one image with scarf and five images without occlusion for each subject in the training set and testing set respectively. The images are cropped to 60×43. Four images on the left in Fig.3 show AR database images with occlusion. In this experiment, we compare our method with a robust version of CRC namely RCRC [18] which considers the robustness to outliers. Table.2 lists the recognition rates of our method and R-CRC on AR database with sunglasses and scarf under different dimensions. One can see that the proposed method is from 2.59% to 12.01% better than R-CRC.

Artificial occlusion in the Extended Yale B database

We use the same training set as section 4.2 then add block occlusion with ratio 30% to the testing set in section 4.2. Block occlusion will cause degradation as the right image shown in Fig.3. The recognition results shown in Table.2 illustrate that our method outperforms RCRC.

5. CONCLUSION

This paper proposed the extension of generalized Rayleigh quotient. It permits to impose suitable constraints. Besides face recognition, this framework can also be applied to other field. For example, the objective function of Canonical Corre-

Table 2. Comparison by using database with occlusion

Dataset	Dimension	60	100	140	180
AR with occlusion	Our method	**0.923**	**0.941**	**0.949**	**0.951**
	RCRC	0.824	0.899	0.916	0.927
/	Dimension	60	70	80	90
Yale B with block occlusion	Our method	**0.683**	**0.700**	**0.712**	**0.721**
	RCRC	0.507	0.577	0.643	0.684

lation Analysis is also in such form. Based on this framework, we impose the graph constraints on collaborative representation residuals to connect samples from same class and distinguish samples from difference classes in low-dimensional space. Our method induces an geometry preserving and discriminant ability increasing projection matrix which improves the performance in the residual-based recognition process. Our experiments on AR, the extended Yale B and face dataset with occlusion prove the advantages of performance and robustness over other state-of-the-art methods.

6. REFERENCES

[1] Rayleigh Lord, *The theory of sound*, Dover Publications, 1945.

[2] Gene H Golub and Charles F Van Loan, *Matrix computations*, vol. 3, JHU Press, 2012.

[3] DJ Mead, "A general theory of harmonic wave propagation in linear periodic systems with multiple coupling," *Journal of Sound and Vibration*, vol. 27, no. 2, pp. 235–260, 1973.

[4] Peter N. Belhumeur, João P Hespanha, and David J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[5] Jian Yang, Alejandro F Frangi, Jing-yu Yang, David Zhang, and Zhong Jin, "Kpca plus lda: a complete kernel fisher discriminant framework for feature extraction and recognition," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 2, pp. 230–244, 2005.

[6] Jian Yang, David Zhang, Jing-yu Yang, and Ben Niu, "Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 4, 2007.

[7] Shuicheng Yan, Dong Xu, Benyu Zhang, Hong-Jiang Zhang, Qiang Yang, and Stephen Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 1, 2007.

[8] Yuchao Liu, Qiang Hua, Xizhao Wang, and Lijie Bai, "Feature extraction using supervised constrained maximum variance mapping," in *Automatic Control and Artificial Intelligence (ACAI 2012), International Conference on*. IET, 2012, pp. 1049–1052.

[9] Jian Yang, Delin Chu, Lei Zhang, Yong Xu, and Jingyu Yang, "Sparse representation classifier steered discriminative projection with applications to face recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 7, pp. 1023–1035, 2013.

[10] John Wright, Yi Ma, Julien Mairal, Guillermo Sapiro, Thomas S Huang, and Shuicheng Yan, "Sparse representation for computer vision and pattern recognition," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031–1044, 2010.

[11] Sheng Huang, Dan Yang, Dong Yang, and Ahmed Elgammal, "Collaborative discriminant locality preserving projections with its application to face recognition," *arXiv preprint arXiv:1312.7469*, 2013.

[12] Fan RK Chung, *Spectral graph theory*, vol. 92, American Mathematical Soc., 1997.

[13] A. M. Martinez, "The ar face database," *Cvc Technical Report*, vol. 24, 1998.

[14] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[15] Lei Zhang, Meng Yang, and Xiangchu Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *Computer vision (ICCV), 2011 IEEE international conference on*. IEEE, 2011, pp. 471–478.

[16] John Wright, Allen Y Yang, Arvind Ganesh, S Shankar Sastry, and Yi Ma, "Robust face recognition via sparse representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[17] Lishan Qiao, Songcan Chen, and Xiaoyang Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognition*, vol. 43, no. 1, pp. 331–341, 2010.

[18] Lei Zhang, Meng Yang, Xiangchu Feng, Yi Ma, and David Zhang, "Collaborative representation based classification for face recognition," *arXiv preprint arXiv:1204.2358*, 2012.