Copyright ⓒ2017 KSII

Noisy label based discriminative least squares regression and its kernel extension for object identification

Zhonghua Liu[1,2*], Gang Liu[1], Jiexin Pu[1], Shigang Liu[3]
[1] Information Engineering College, Henan University of Science and Technology
Luoyang, 471023 - China
[e-mail: lzhua_217@163.com]
[2] Fujian Provincial Key Laboratory of Information Processing and Intelligent Control , Minjiang University
FuZhou, 350108 - China
[3]School of Computer Science, Shaanxi Normal University
Xi'an, 710062 - China
*Corresponding author: Zhonghua Liu

Received October 2, 2016; revised December 22, 2016; revised February 5, 2017; accepted February 27, 2017; published May 31, 2017

Abstract

In most of the existing literature, the definition of the class label has the following characteristics. First, the class label of the samples from the same object has an absolutely fixed value. Second, the difference between class labels of the samples from different objects should be maximized. However, the appearance of a face varies greatly due to the variations of the illumination, pose, and expression. Therefore, the previous definition of class label is not quite reasonable. Inspired by discriminative least squares regression algorithm (DLSR), a noisy label based discriminative least squares regression algorithm (NLDLSR) is presented in this paper. In our algorithm, the maximization difference between the class labels of the samples from different objects should be satisfied. Meanwhile, the class label of the different samples from the same object is allowed to have small difference, which is consistent with the fact that the different samples from the same object have some differences. In addition, the proposed NLDLSR is expanded to the kernel space, and we further propose a novel kernel noisy label based discriminative least squares regression algorithm (KNLDLSR). A large number of experiments show that our proposed algorithms can achieve very good performance.

Keywords: Noisy label; discriminative least squares regression (DLSR); noisy label based discriminative least squares regression algorithm (NLDLSR); pattern recognition

https://doi.org/10.3837/tiis.2017.05.012 **ISSN : 1976-7277**

1. Introduction

In the past few decades, feature extraction has been widely studied. Many feature extraction algorithms were presented in the literatures and they are extensively applied to the image analysis, biometric recognition [1-4], and multi-graph learning [5-7]. Nowadays, the existing recognition methods can achieve very good recognition effect under restricted conditions, such as standard illumination, normal expression and fixed pose [8-10]. However, these factors such as illumination, expression and pose are generally uncontrolled in practical application, which can cause the recognition performance of the algorithms to drop sharply. Therefore, Face recognition is still a challenging task, and we should put more effort to research and solve it.

It is well known that least squares regression (LSR) is a widely-used statistical analysis technique. Recently, LSR is used in the field of image recognition. By evaluating linear representation ability of training samples for any test sample, LSR can solve image classification problems. Because LSR is very simple and highly effective for image classification and recognition, a huge number of variant forms of LSR have been presented. By introducing sparse constraint to least squares regression (LSR), Wright et al. [11] proposed the sparse representation based classification (SRC) algorithm. In SRC, any test sample is firstly represented as a linear combination of all training samples. Then, by solving the problem of sparse optimization, the sparse representation coefficients can be obtained. Lastly, the test sample is assigned to the class in which there is minimum residual between the test sample and the corresponding reconstruction sample of each class. SRC can achieve good recognition performance in image recognition. To expand SRC to the kernel space, Yin et al. [12] proposed a kernel sparse representation based classification (KSRC) algorithm. In order to eliminate the influence of occlusions or illumination variations in image recognition, a manifold regularized local sparse representation (MRLSR) algorithm [13] is presented. In MRLSR, both local similarity and individual sparsity are simultaneously obtained. Based on Gabor transform, the Gabor-feature based SRC (GSRC) algorithm was proposed [14], in which the computational cost can be significantly reduced in coding the occluded face images. He et al. [15] presented the correntropy-based sparse representation (CSR) [15] algorithm. In CSR, the maximum correntropy criterion is used. CSR is very different from SRC, and the difference is as follows. It can deal with non-Gaussian noise and outliers [16]. Yuan et al. [17] proposed a multi-task joint sparse representation based classification (MTJSRC) algorithm. MTJSRC can compute joint sparse representation of visual signal across multiple kernel-based representations. L_1 norm based linear representation methods can get better performance in pattern recognition and image classification. However, it need to take more time to solve L_1 norm optimization problem, and it can not meet real-time requirements. The low efficiency of L_1 norm based methods is largely because of L_1 norm based optimization problem. We know that it is very time consuming to compute L_1 optimization problem. After the working mechanism of SRC is analyzed and researched, Zhang et al. [18] gave their own viewpoint. In SRC algorithm, it is the collaborative representation, but not the L1-norm sparsity that plays a key role for image classification [18]. Based on this observation, they proposed a collaborative representation based classification (CRC) method [18]. For any object, there are the following facts. Firstly, each object usually includes many different features. Secondly, each different feature should have different contributions to the pattern classification and representation [19]. Based on this fact, Yang et al [20] presented a relaxed collaborative representation (RCR) algorithm. Both the distinctiveness of features and the similarity of features can be effectively utilized in RCR. Xu et al. [21] proposed a two-phase test sample sparse representation (TPTSSR) algorithm. In TPTSSR, a coarse-to-fine classification strategy is used for classification and recognition, and it can achieve good classification performance. Any sample from an object usually lies in a linear subspace constructed by the samples from the same object [22]. Based on the fact, Togneri et al. [23] presented linear regression classification (LRC) algorithm. The variation of face pose can lead to the sharp degradation of the classification performance. In order to solve the problem, Chai et al. [24] presented a local linear regression (LLR) algorithm, in which the virtual frontal view corresponding to each non-frontal training face image can be generated. By modifying the classification rule of the minimum squared error (MSE), Xu et al. [25] proposed a modified minimum squared error classification (MMSEC) algorithm. In MMSEC, the class label of the test sample is firstly predicted, and the training samples nearest to it are simultaneously obtained. Then the obtained predicted results are used to classify the test sample. However, the statistical information from the training samples, which actually plays a major role in classification and recognition, are usually ignored in most of the linear representation based methods. In order to take full advantage of the statistical information, Qian et al. [26] proposed a general regression and representation (GRR) algorithm for image recognition and classification. By using a simple and effective linear regression function to map the new data points, Nie et al. [27] proposed a novel flexible manifold embedding (FME) algorithm. Both the label information from labeled data and a manifold structure from all data can be effectively used in FME.

The previous works largely focused on investigating the importance of the L1-regularizer/L2-regularizer. The class label information is almost not emphasized. In least squares regression (LSR) and its variants, the class label is usually defined as follows. Firstly, the samples from the same object have the same fixed class label. Secondly, the difference of class label corresponding to different objects should be maximized. However, human face is non-rigid, and the variations of pose, illumination and expression can lead to the large variation of face images even for the same object. Due to the diversity of face images, the definition of the above class label is not quite reasonable.

In order to enlarge the distance between different classes under the framework of LSR, a discriminative least squares regression (DLSR) algorithm is proposed [28]. In DLSR, the -dragging strategy can force the regression targets of different classes moving along opposite directions, which can make the distances between classes enlarge. However, the class label of the samples from the same object is almost the same. Therefore, the class label defined in DLSR is essentially the same as the above defined. However, inspired by the -dragging idea, we propose a noise-label based discriminative least squares regression (NLDLSR) algorithm in this paper. In the proposed NLDLSR algorithm, the class label is defined as follows. First, the distance among class labels corresponding to different objects should be enlarged. Second, the distance among the class labels of different samples from the same object is allowed to have small difference, which is consistent with the fact that the different samples from the same object have a certain difference. In the proposed NLDLSR algorithm, the definition of class label is more general, which can help to obtain the robust mapping relationship between training samples and their class label in training stage. Therefore, the proposed NLDLSR algorithm is simple but effective.

The remaining parts of this paper are organized as follows. We briefly introduce the discriminative least squares regression (DLSR) algorithm in section 2. Section 3 describes the proposed noise-label based discriminative least squares regression (NLDLSR) algorithm in detail. In section 4, Based on the proposed NLDLSR, we present a novel kernel noise-label based discriminative least squares regression (KNLDLSR) algorithm. The experimental results are given in section 5. Section 6 gives our conclusion.

2. Discriminative least squares regression (DLSR) algorithm

Supposed that we have n training samples $\{(x_i, y_i)\}_{i=1}^{n}$ from $c\ (c \geq 2)$ classes, where x_i is a data point in R^m and $y_i \in \{1, 2, \cdots, c\}$ is the class label corresponding to the data x_i. The linear equation can be satisfied as follows [28].

$$XW + e_n t^T \approx Y \tag{1}$$

where $X = [x_1, x_2, \cdots, x_n]^T \in R^{n\times m}$, $Y = [f_{y_1}, f_{y_2}, \cdots, f_{y_n}]^T \in R^{n\times c}$. For the i^th^ ($i = 1, 2, \cdots, c$) class, $f_i = \left[\underbrace{0,\cdots,0}_{\text{i-1}},1,0,\cdots,0\right]^{\mathrm{T}} \in R^c$, W is a transformation matrix in $R^{m\times c}$, t is a translation vector in R^c, $e_n = [1, 1, \cdots, 1]^T \in R^n$ is a vector with all 1s.

The i^{th} column vector of matrix Y is actually assigned to a type of binary regression with target "+1" for the i^{th} class and target "0" for the rest of class. In order to get good classification performance, the binary outputs are requested to be far away along two opposite directions. Namely, the output will be $1 + \varepsilon_i$ for the samples from one object and $-\varepsilon_i$ for the samples from other objects. where ε_i is a small positive variable [28].

Supposed that $B \in R^{n\times c}$ be a constant matrix. An element B_{ij} from the i^{th} row and j^{th} column of B is defined as follows.

$$B_{ij} = \begin{cases} +1, & if\ y_i = j \\ -1, & otherwise \end{cases} \tag{2}$$

Each element in matrix B corresponds to a dragging direction. The ε-dragging strategy is used on each element of Y, and the matrix $M \in R^{n\times c}$ records these $\{\varepsilon\}$ [28]. The residual can be gotten as follows.

$$XW + e_n t^T - (Y + B \otimes M) = 0 \tag{3}$$

where $\otimes$ is a Hadamard product operator of matrices.

The objective function of DLSR algorithm is defined as follows [28].

$$\begin{aligned} &\min_{W,t,M} \left\| XW + e_n t^T - Y - B \otimes M \right\|_F^2 + \lambda \|W\|_F^2 \\ &s.t.\ \ M \geq 0 \end{aligned} \tag{4}$$

where λ is a positive regularization parameter. By solving the optimization problem of Equation (4), we can obtain the optimal W and t as follows [28].

$$W = (X^T HX + \lambda I_m)^{-1} X^T HR \tag{5}$$

$$t = \frac{(R^T e_n - W^T X^T e_n)}{n} \tag{6}$$

where I_m is a $m \times m$ identity matrix, $R = Y + B \otimes M$, and $H = I_n - (1/n)e_n e_n^T$, in which I_n is a $n \times n$ identity matrix.

3. Noisy-label based discriminative least squares regression

Deformable objects usually have a variety of different appearance. For example, face appearance can vary dramatically due to the variations of poses, expressions and illuminations. The variations between the images corresponding to the same person due to poses, expressions and illuminations are almost always greater than the variations due to changes in face identity [29, 30]. That is to say, even if the face images come from the same person, the difference among the images may be very large. Taking into account the fact, we think that the class label of different images from the same object is not necessary to be exactly the same. It should be permitted to have minor differences. The proposed NLDLSR algorithm can obtain the optimal mapping relationship between the training samples and their corresponding class labels during the training stage.

Based on the above idea, the Equation (1) can be rewritten as

$$XW + e_n t^T = Y^{'} \tag{7}$$

where $Y^{'} = [f_{y_1}^{'}, f_{y_2}^{'}, \cdots, f_{y_n}^{'}]^T \in R^{n \times c}$, and $f_j^{'} = \left[\underbrace{0, \cdots 0}_{\text{j-1}}, 1+\sigma, 0, \cdots, 0 \right]^{\mathrm{T}} \in R^c$. In order to truly reflect the difference of class label from the same object, σ is randomly generated, and it is a small real-number variable.

The objective function of the proposed NLDLSR is defined as follows.

$$\min_{W,t} \left\| XW + e_n t^T - Y^{'} \right\|_F^2 + \lambda \left\| W \right\|_F^2 \tag{8}$$

Based on convex optimization theory, the optimal solution of Equation 8 can be easily obtained as follows.

$$W = (X^T HX + \lambda I_m)^{-1} X^T HY^{'} \tag{9}$$

$$t = \frac{(Y^{'} e_n - W^T X^T e_n)}{n} \tag{10}$$

The specific steps of the proposed NLDLSR are summarized as follows.

Step 1. The columns of training samples X are normalized to have unit L_2-norm.

Step 2. Let y be the test sample. Use $y = \frac{y}{\| y \|}$ converts the test sample y into unit vectors.

Step 3. Solving the optimization model of Equation 8, and W and t can be gotten as follows.

$W = (X^T HX + \lambda I_m)^{-1} X^T HY$

$t = (Y^T e_n - W^T X^T e_n) / n$

Step 4. The class label $ylabel \in R^{c \times 1}$ of the test sample y can be obtained by the following Equation

$ylabel = W^T y + t$

Step 5. Respectively compute the "distance" between the class label $ylabel$ and each standard class label f_i $(i=1,2,\cdots,c)$ from i^th class.

$$dis_i = \|ylabel - f_i\|_2$$

Step 6. The identity of test sample y is output as

$$identity(y) = \arg\min\{dis_i\}$$

4. Kernel noise-label based discriminative least squares regression

The proposed NLDLSR algorithm is extended to kernel space, and we propose a novel kernel noise-label based discriminative least squares regression (KNLDLSR) algorithm. KNLDLSR algorithm is assumed to be the nonlinear kernel version of the proposed NLDLSR by means of kernel method [31, 32], which mainly deal with the classification issue of nonlinear feature.

Assume that φ be a nonlinear function. Each training sample from low dimensional space can be converted to a high dimensional space by the nonlinear function φ .The corresponding training samples in high dimensional space are respectively denoted by $\varphi(x_1),\varphi(x_2),\cdots,\varphi(x_n)$. In KNLDLSR algorithm, the linear Equation can be rewritten as follows.

$$\varphi(x)\tilde{W} + e_n t^T = Y' \tag{11}$$

where $\varphi(x) = [\varphi(x_1),\varphi(x_2),\cdots,\varphi(x_n)]^T$, $\tilde{W}$ is a transform matrix, and e_n, t, Y' have the same meaning as the corresponding variables in the above definition. It can be known from the kernel theory that matrix $\tilde{W}$ in Equation (11) may be represented as a linear combination of all training samples $\varphi(x_i), i=1,2,\cdots,n$. Namely,

$$\tilde{W} = \sum_{i=1}^{n} \beta_i \varphi(x_i) \tag{12}$$

where $\beta = [\beta_1, \beta_2, \cdots \beta_n]^T$. Thus, Equation (11) can be rewritten as

$$\begin{aligned} &\varphi(x)\sum_{i=1}^{n}\beta_i\varphi(x_i) + e_n t^T = Y' \\ &\Rightarrow \varphi(x)\varphi(x)^T\beta + e_n t^T = Y' \end{aligned} \tag{13}$$

A kernel function can be used to represent the dot product in the Hilbert space, such as $K(x_i, x_j) = \varphi(x)\varphi(x)^T$. Equation (13) can be rewritten as

$$K\beta + e_n t^T = Y' \tag{14}$$

Therefore, the objective function of the proposed KNLDLSR algorithm is defined as follows.

$$\min_{\beta,t} \left\|K\beta + e_n t^T - Y'\right\|_F^2 + \lambda\|\beta\|_F^2 \tag{15}$$

From Equation (15), a new function $g(\beta,t)$ can be denoted by $g(\beta,t) = \left\| \left\|K\beta + e_n t^T - Y'\right\|_F^2 + \lambda\|\beta\|_F^2 \right\|$. According to the relevant optimization theory of the matrix, we can get

$$\begin{aligned}&\frac{\partial g(\beta,t)}{\partial t}=0\\&\Rightarrow \beta^T K^T e_n + t e_n^T e_n - (Y^{'})^T e_n = 0\\&\Rightarrow t = ((Y^{'})^T e_n - \beta^T K^T e_n)/n\end{aligned} \tag{16}$$

$$\begin{aligned}&\frac{\partial g(\beta,t)}{\partial \beta}=0\\&\Rightarrow K^T(K\beta + e_n e_n^T Y^{'}/n - e_n e_n^T K\beta/n - Y^{'}) + \lambda\beta = 0\\&\Rightarrow K^T(I_n - e_n e_n^T/n)K\beta - K^T(I_n - e_n e_n^T/n)Y^{'} + \lambda\beta = 0\\&\Rightarrow \beta = (K^T H K + \lambda I_m)^{-1} K^T H Y^{'}\end{aligned} \tag{17}$$

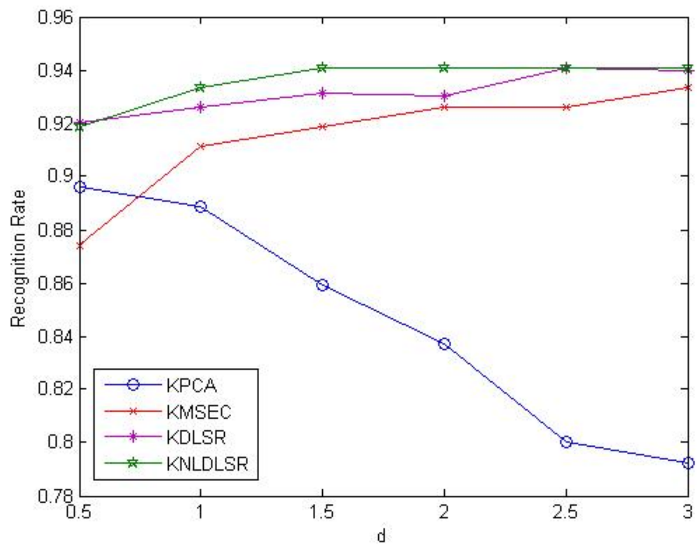
where I_m, I_n and H is the same meaning as defined in above Equations.

5. Experiments

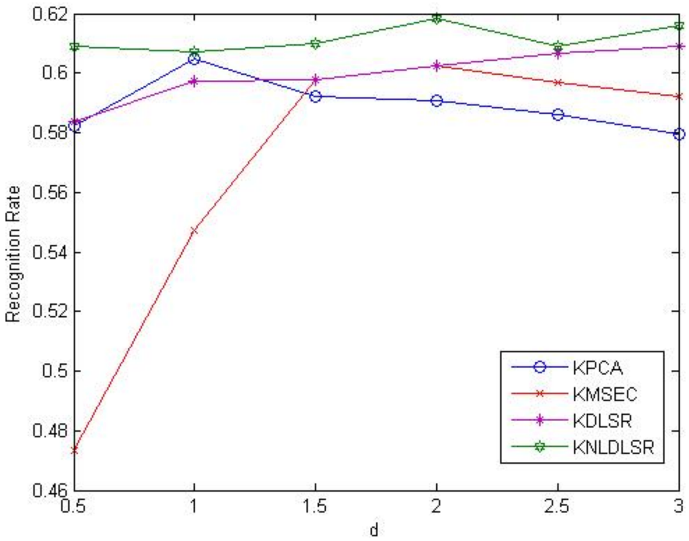
In this section, we use four public face databases to validate the proposed NLDLSR and KNLDLSR algorithms. When the proposed NLDLSR is run, PCA[33], MSEC and DLSR algorithms are also respectively performed for comparisons [28]. When the proposed KNLDLSR is performed, the KPCA[34], KMSEC[35], and KDLSR are used to act as comparative algorithms. In addition, the polynomial kernel $k(x,y)=(1+x^T y)^d$ is used in the proposed KNLDLSR and the corresponding kernel algorithms. In the algorithms, there are two parameters σ and d. The optimal σ is selected for the proposed NLDLSR algorithm, and we make $|\sigma| \le 0.1$ in this paper. The optimal kernel parameter d is respectively selected for each kernel algorithms in experiments.

5.1 Parameters Selection

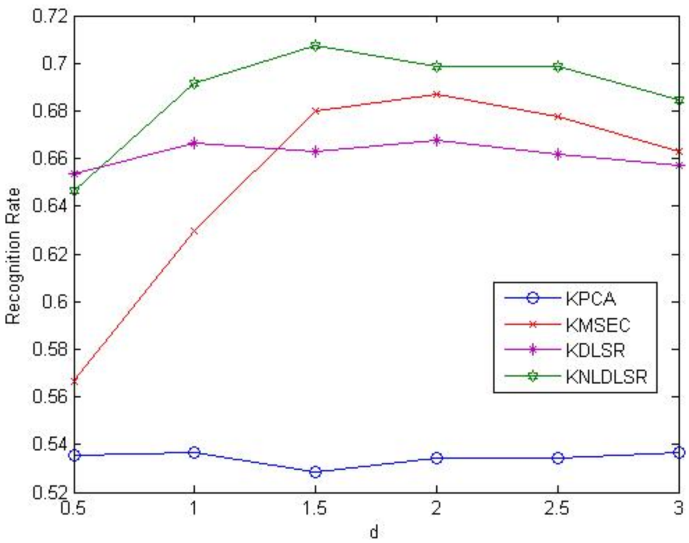
In the proposed KNLDLSR method, there is only one parameter d. In order to determine the optimal value of the parameter d, we do some experiments on public databases. The average recognition rates of KNLDLSR vary with the parameter d in **Fig. 1**. It can be seen from **Fig. 1** that the value of the parameter d will vary on different face databases.



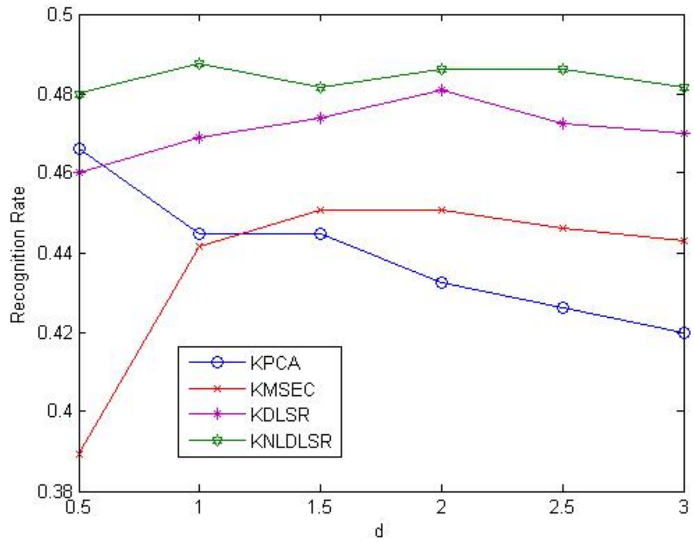
(a)


0.62
0.6
0.58
0.56
0.54
0.52
0.5
0.48
0.46
Recognition Rate
0.5
1
1.5
2
2.5
3
d
KPCA
KMSEC
KDLSR
KNLDLSR

(b)


0.72
0.7
0.68
0.66
0.64
0.62
0.6
0.58
0.56
0.54
0.52
Recognition Rate
0.5
1
1.5
2
2.5
3
d
KPCA
KMSEC
KDLSR
KNLDLSR

(c)



(d)

Fig. 1. The performance of KNLDLSR versus parameters d. (a) Yale, (b) UMIST, (c) AR, (d)GT

5.2 Experiment on Yale face database

The Yale face database includes 165 images from 15 objects, and each object has 11 images under various lighting conditions and facial expressions. In this experiment, we manually crop each image and make them to 50×40 pixels. Some sample images of one person are shown in **Fig. 2.**

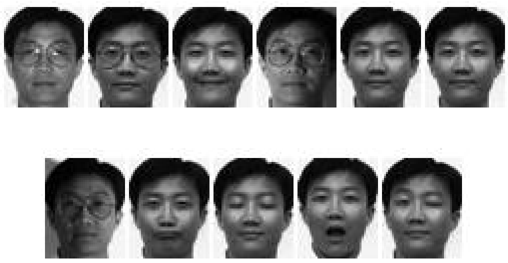


Fig. 2. Sample face images from the Yale database.

In this section, the first 1, 2 up to 10 face images of each object are respectively selected to serve as the training set and the rest images of each object to act as test set. The class label of each training sample is not fixed in our proposed algorithms, and it is randomly generated. In order to truly reflect the performance of our algorithms, we run the proposed algorithms for 10 times to eliminate the effect of random for our algorithms. The **Table 1** shows the recognition results.

Table 1. Recognition rates of different methods on the Yale database (%)

Methods	Number of the original training samples per class									
	1	2	3	4	5	6	7	8	9	10
PCA	74	88.89	90	92.38	88.89	86.67	98.33	100	100	100
MSEC	88.00	92.59	95.83	100	100	100	100	100	100	100
DLSR	88.67	92.59	95.83	100	100	100	100	100	100	100
NLDLSR	88.80	93.48	95.33	100	100	100	100	100	100	100
KPCA(d=0.6)	76.67	88.89	90.00	92.38	88.89	88.00	98.33	100	100	100
KMSEC(d=2.5)	88.00	92.59	95.83	100	100	100	100	100	100	100
KDLSR(d=2.5)	89.33	94.07	95.83	100	100	100	100	100	100	100
KNLDLSR(d=2.5)	90	94.81	95.83	100	100	100	100	100	100	100

It can be seen from **Table 1** that we can draw two conclusions. Firstly, the proposed NLDLSR outperforms PCA, MSEC and DLSR algorithms. Secondly, the proposed KNLDLSR achieves the highest recognition rate in all algorithms. From the above two points we can know that the kernel method can indeed enhance the classification performance and improve the recognition rate.

5.3 Experiments on UMIST face database

The UMIST database contains 575 face images from 20 persons. The individuals are a mix of appearance, race, and sex, and are captured in a range of poses from profile to frontal views. The number of different views per person varies from 19 to 48. The size of each face image is manually cropped to 56×48 pixels. **Fig. 3** shows some sample images from one object.

Fig. 3. Sample face images from the UMIST database

In this experiment, the first 2, 3 to 7 face images of each object are respectively selected to act as training set and the rest of each person to take as test images. The proposed algorithms are performed for 10 times. **Table 2** lists the recognition results.

Table 2. Recognition rates of different methods on the UMIST database (%)

Methods	Number of the original training samples per class					
	2	3	4	5	6	7
PCA	60.00	71.72	75.63	84.60	88.51	89.66
MSEC	58.28	69.79	75.31	83.29	88.74	90.57
DLSR	60.00	74.25	80.23	85.98	88.74	91.95
NLDLSR	60.69	73.79	81.61	88.28	89.89	93.10
KPCA(d=1.0)	60.46	72.41	78.16	85.75	88.97	90.57
KMSEC(d=1.7)	60.46	72.87	77.70	84.83	87.13	90.57
KDLSR(d=3.5)	60.92	74.71	79.77	86.44	89.20	92.64
KNLDLSR(d=2.0)	61.84	74.48	81.15	88.89	89.91	93.56

From **Table 2**, we can see that the recognition rates of proposed NLDLSR are slightly higher than those of the other algorithms when the training sample size is 2. Though the recognition performance of the proposed KNLDLSR is better than other kernel methods, the performance of the proposed NLDLSR and KNLDLSR is relatively close. This is because the parameter d in polynomial kernel function is set as a fixed value regardless of the training sample size.

5.4 Experiment on AR face database

The AR face database [36] consists of more than 4000 color face images from 126 persons, including 26 frontal views of faces with different occlusions, facial expressions, and lighting conditions for each person. The images from 120 persons were taken in two sessions (14 days apart) and each session includes 13 color images. Fourteen face images (each session including 7) of these 120 persons are used in our experiment. The images are firstly converted to grayscale. The size of each face image is manually cropped to 50×40 pixels. Some images of one person are shown in **Fig. 4**.

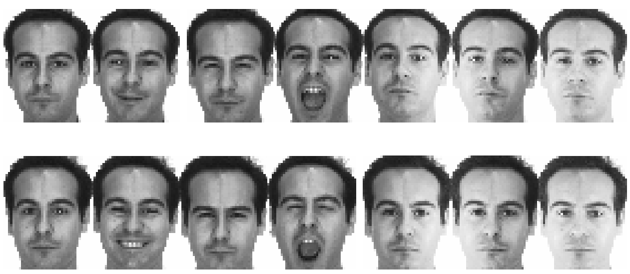


Fig. 4. Sample face images from AR database.

For this experiment, fourteen not occluded face images of each object from the first session and the second session are used to validate our algorithms. We respectively select The l face images (l varies from 1 to 7) are respectively selected from the first session to take as training images, and the seven face images of each object from the second session are selected to act as test images. The proposed algorithms are repeated for 10 times. **Table 3** shows the experiment results.

Table 3. Recognition rates of different methods on the AR database (%)

Methods	Number of the original training samples per class						
	1	2	3	4	5	6	7
PCA	43.93	43.45	46.07	46.31	53.69	60.36	62.98
MSEC	67.02	67.14	66.79	65.83	70.06	74.52	75.12
DLSR	68.21	67.38	67.62	66.07	70.60	75.48	75.71
NLDLSR	68.89	69.17	70.71	70.60	73.10	76.55	77.38
KPCA(d=0.7)	52.74	53.75	55.12	55.36	60.00	64.40	67.26
KMSEC(d=2.0)	68.57	68.69	69.76	69.29	71.19	76.31	77.14
KDLSR(d=1.7)	66.07	66.43	67.74	68.69	70.12	76.55	77.74
KNLDLSR(d=1.5)	69.40	70.95	71.07	70.60	72.74	77.62	78.33

From **Table 3**, we can know that the proposed NLDLSR and KNLDLSR can respectively achieve the best recognition performance in all algorithms (the corresponding linear methods and the corresponding kernel methods) irrespective of the variation of training sample size. We also find that the kernel versions of the algorithms do not consistently outperform the original algorithms. The reasons may be as follows. The parameter d in polynomial kernel function is very important, and its value can seriously affect the performance of the algorithms. In order to get the best recognition performance for kernel methods, the parameter d usually should be selected different value when the training sample size changes. However, for simplicity, we set the parameter d fixed value irrespective of the variation of training sample size in our experiments.

5.5 Experiments on Georgia Tech face database

Georgia Tech face database includes images from 50 persons taken in two or three sessions, which were built at Georgia Institute of Technology. For each object in the database, 15 color images which are captured under cluttered background were obtained at the resolution of 640×480 pixels. Faces shown in these images may be tilted and/or frontal with different expressions, scales, and illuminations. All images were resized to 60×50 pixels. All color images are converted to grayscale images. **Fig. 5** shows some images of one person.

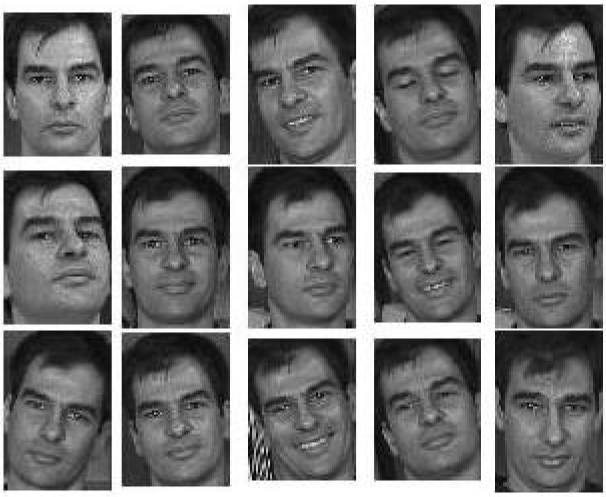


Fig. 5. Sample face images from Georgia Tech face database.

In this experiment, the first 1, 2 up to 7 face images of each person are respectively used to take as training set and the rest to be test set. The proposed algorithms are repeated for 10 times. The recognition results of all algorithms are shown as **Table 4**.

Table 4. Recognition rates of different methods on the Georgia Tech database (%)

Methods	Number of the original training samples per class				
	1	2	3	4	5
PCA	36.57	45.23	47.67	51.82	53.60
MSEC	33.00	42.77	46.00	48.36	51.60
DLSR	33.71	44.46	49.00	50.55	54.40
NLDLSR	36.14	45.38	49.00	53.64	56.80
KPCA(d=0.7)	38.43	45.08	48.67	50.36	52.20
KMSEC(d=2.0)	34.29	45.08	49.17	53.27	56.40
KDLSR(d=1.7)	35.43	48.15	50.33	54.55	57.20
KNLDLSR(d=1.5)	35.29	48.62	51.83	54.55	59.00

It can be seen from **Table 4** that the proposed NLDLSR and KNLDLSR algorithms respectively outperform other corresponding algorithms when training sample size is respectively 2, 3, 4 and 5. When the training sample size is one, PCA and KPCA respectively get the highest recognition rates in all algorithms.

From all above experiments, we can see that the proposed NLDLSR indeed improve the classification performance. Therefore, it is reasonable to have small difference for class label from the same object.

6. Conclusion

In this paper, we propose a novel noisy-label based discriminative least squares regression (NLDLSR) algorithm for multi-class classification. Compared with the discriminative least squares regression (DLSR) algorithm, the main contributions of our method are as follows. 1) The optimal mapping relationship between the training samples and the corresponding class label can be obtained during the training stage. 2) The optimization problem in DLSR is solved in an iterative manner, and training time is relatively long. However, the optimal solution of the proposed NLDLSR algorithm can be obtained in one step, so the time complexity in training stage is very short. Moreover, the proposed NLDLSR algorithm is extended to the kernel space and we present a KNLDLSR algorithm. Extensive experiments show that the proposed methods have high classification accuracy.

References

[1] Y. Xu, A. Zhong, et al., "LPP solution schemes for use with face recognition," *Pattern Recognition*, 43(12):4165-4176, 2010. Article(CrossRef Link)

[2] Wankou Yang, Zhenyu Wang, Changyin Sun., "A collaborative representation based projections method for feature extraction," *Pattern Recognition*, 48(1): 20-27, 2015. Article(CrossRef Link)

[3] Z. Lai, Y. Xu, et al, "Multilinear sparse principal component analysis," *IEEE Transactions on Neural Networks and learning Systems*,25(10): 1942-1950, 2014. Article(CrossRef Link)

[4] Lingfen Wang, Huaiyu Wu, "Chunhong Pan. Manifold regularized local sparse representation for face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, 25(4): 651-659, 2015. Article (CrossRef Link)

[5] J. Wu, S. Pan, X. Zhu, C. Zhang, and X, Wu, "Positive and Unlabeled Multi-Graph Learning," *IEEE Trans. Cybern*, 2016. Article (CrossRef Link)

[6] J. Wu, S. Pan, X. Zhu, Z. Cai, "Boosting for Multi-Graph Classification," *IEEE Trans. Cybern*, 45(3): 416-429, 2015. Article (CrossRef Link)

[7] J. Wu, X. Zhu, C. Zhang, P. S. Yu, "Bag Constrained Structure Pattern Mining for Multi-Graph Classification," *IEEE Trans. Knowl. Data Eng.*, 26(10): 2382-2396, 2014. Article (CrossRef Link)

[8] Fujiao Ju, Yangfeng Sun, et al, "Image outlier detection and feature extraction via L1-norm-based 2D probabilistic PCA," *IEEE Transactions on Image Processing*, 24(12): 4834-4846, 2015. Article (CrossRef Link)

[9] Yong Xu, Xingjie Zhu, et al, "Using the original and 'symmetrical face training samples to perform representation based two-step face recognition," *Pattern Recognition*, 46: 1151-1158, 2013. Article (CrossRef Link)

[10] Shenghua Gao, Yuting Zhang, et al, "Single sample face recognition via learning deep supervised autoencoders," *IEEE Transactions on Information Forensics and Security*, 10(10): 2108-2118,2015. Article (CrossRef Link)

[11] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2): 210-227, 2009. Article (CrossRef Link)

[12] Jun Yin, Zhonghua Liu, et al, "Kernel sparse representation based classification," *Neurocomputing*, 77(11): 120-128, 2012. Article (CrossRef Link)

[13] Lingfeng Wang, Huaiyu Wu, Chunhong Pan, "Manifold regularized local sparse representation for face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, 25(4): 651-659, 2015. Article (CrossRef Link)
[14] M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary," in *Proc. of 11th Eur. Conf. Comput. Vis. (ECCV)*, pp. 448–461, 2010. Article (CrossRef Link)
[15] Ran He, Weishi Zheng, Baogang Hu, "Maximum correntropy criterion for robust face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8): 1561-1576, 2011. Article (CrossRef Link)
[16] W. Liu, P.P. Pokharel, and J.C. Principe, "Correntropy: Properties and Applications in Non-Gaussian Signal Processing," *IEEE Trans. Signal Processing*, 55(11): 5286-5298, 2007. Article (CrossRef Link)
[17] X. Yuan and S. Yan, "Visual classification with multi-task joint sparse representation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 3493–3500, Jun. 2010. Article (CrossRef Link)
[18] L. Zhang, Meng Yang, Xiangchu Feng, "Sparse Representation or Collaborative Representation: Which Helps Face Recognition?" in *ICCV* 2011, pp.1-8, 2011. Article (CrossRef Link)
[19] J. Wu, S. Pan, X. Zhu, P. Zhang, and C. Zhang, "Self-Adaptive One-Dependence Estimators for classification," *Pattern Recognit.*, 51: 358-377, 2016. Article (CrossRef Link)
[20] Meng Yang, Lei Zhang, et al, "Relaxed collaborative representation for pattern classification," in *Proc. of International Conference on Wavelet Analysis and Pattern Recognition*, pp.13-16, 2014. Article (CrossRef Link)
[21] Yong Xu, David Zhang, et al., "A two-phase test sample sparse representation method for use with face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, 21(9): 1255-1262, 2011. Article (CrossRef Link)
[22] R. Barsi and D. Jacobs, "Lambertian Reflection and Linear Subspaces. IEEE Trans," *Pattern Analysis and Machine Intelligence*, 25(2): 218-233, 2003. Article (CrossRef Link)
[23] Roberto Togneri, Mohammed Bennamoun, "Linear regression for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11): 2106-2112, 2010. Article (CrossRef Link)
[24] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally Linear Regression for Pose-Invariant Face Recognition," *IEEE Trans. Image Processing*, vol. 16, no. 7, pp. 1716-1725, July 2007. Article (CrossRef Link)
[25] Yong Xu, Xiaozhao Fang, et al, "Modified minimum squared error algorithm for robust classification and face recognition experiments," *Neurocomputing*, 135(5): 253-261, 2014. Article (CrossRef Link)
[26] Jianjun Qian, Jian Yang, "General regression and representation model for face recognition," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp.166-172, 2013. Article (CrossRef Link)
[27] Feiping Nie, Dong Xu, et al, "Flexible manifold embedding: A framework for semi-supervised and unsupervised dimension reduction," *IEEE Transactions on Image Processing*, 19(7): 1921-1932, 2010. Article (CrossRef Link)
[28] Shiming Xiang, Feiping Nie, et al, "Discriminative least squares regression for multiclass classification and feature selection," *IEEE Transactions on Neural Networks and Learning Systems*, 23(11): 1738-1754, 2012. Article (CrossRef Link)
[29] X. Cao, W. Shen, et al, "Illumination invariant extraction for face recognition using neighboring wavelet coefficients," *Pattern Recognition*, 45: 1299-1305, 2012. Article (CrossRef Link)
[30] Amirhosein Nabatchian, Esam Abdel-Raheem, Majid Ahmadi, "Illumination invariant feature extraction and mutual-information-based local matching for face recognition under illumination variation and occlusion," *Pattern Recognition*, 44: 2576-2587, 2011. Article (CrossRef Link)
[31] Meng Yang, Lei Zhang, et al, "Robust kernel representation with statistical local features for face recognition," *IEEE Transactions on Neural Networks and Learning Systems*, 24(6): 900-912, 2013. Article (CrossRef Link)

[32] Cuicui Kang, Shengcai Liao, et al, “Kernel sparse representation with pixel-level and region-level local feature kernels for face recognition,” *Neurocomputing*, 133(10): 141-152, 2014. Article (CrossRef Link)

[33] M.A. Turk, A.P. Pentland, “Face recognition using eigenfaces,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586-591, 1991. Article (CrossRef Link)

[34] B. Scholkopf, A. Smola, K. R. Muller, “Nonlinear component analysis as a kernel eigenvalue problem,” *Neural computation*, 10:1299-1319, 1998. Article (CrossRef Link)

[35] Yong Xu, “A new kernel MSE algorithm for constructing efficient classification procedure,” *International Journal of Innovative Computing, Information and Control*, 5(8): 2439-2447, 2009. Article (CrossRef Link)

[36] A. Martinevz, and R. benavente, “The AR face database,” *CVC Tech. Report*, No. 24, 1998. Article(CrossFef Link)

Zhonghua Liu received his B.S. in computer engineering from the First Aeronautical Institute of the Air Force , his M.S. in computer software and theory from Xihua University, and his PhD in pattern recognition and intelligentsystems from Nanjing University of Science and Technology in 1998, 2005 and 2011, respectively. He is an associate professor in the Information Engineering College, Henan University of Science and Technology.His current research interests include mattern recognition, face recognition, image processing and scene matching.

Gang Liu received the Ph.D. degree from Northwestern Polytechnical University, P. R. China. His research interests include image processing, multi-scale methods and artificial intelligence.

Jiexin Pu is a professor in the Information Engineering College, Henan University of Science and Technology.His research interests include computer application technology, pattern recognition and Computer network and information security.

Shigang Liu is a professor in School of Computer Science, Shaanxi Normal University Xi'an, China. His research interests include pattern recognition, image processing and artificial intelligence.